(19) Europäisch s Pat ntamt
Europ an Pat nt Offic
Offic urop´en des brevets

(11) EP 1 318 195 A1

(12) EUROPEAN PATENT APPLICATION

(43) Date of publication:
11.06.2003 Bulletin 2003/24

(21) Application number: **01204762.7**

(22) Date of filing: **10.12.2001**

(51) Int Cl.[7]: **C12N 15/12**, C07K 16/00, G01N 33/50, C12N 5/10, C07K 1/14, A61K 39/395, A61K 47/48, C12N 15/10

(84) Designated Contracting States:
AT BE CH CY DE DK ES FI FR GB GR IE IT LI LU MC NL PT SE TR
Designated Extension States:
AL LT LV MK RO SI

(71) Applicant: **CatchMabs B.V.**
6708 PW Wageningen (NL)

(72) Inventors:
- **Houtzager, Erwin**
  3958 XD, Amerongen (NL)
- **Vijn, Irma Maria Caecilia Vijn**
  6721 ZJ, Bennekom (NL)
- **Sijmons, Peter Christiaan**
  1075 GK, Amsterdam (NL)

(74) Representative: **Prins, Adrianus Willem et al**
Vereenigde,
Nieuwe Parklaan 97
2587 BN Den Haag (NL)

Remarks:
The sequence listing, which is published as annex to the application documents, was filed after the date of filing. The applicant has declared that it does not include matter which goes beyond the content of the application as filed.

(54) A structure for presenting desired peptide sequences

(57) The binding properties of proteinaceous molecules to their targets was found to be manipulatable not only at the level of the sequence of the binding part on the proteinaceous molecule but also at the level of the structure of the proteinaceous molecules itself. Core sequences are identified that can be used and manipulated to obtain the desired binding properties of a binding peptide. The present invention further provides means and methods for selecting and designing such cores and binding peptides. The invention further provides uses of such proteinaceous molecules.

EP 1 318 195 A1

D scription

[0001] The invention relates to methods and means for providing binding molecules with improved properties, be it in binding or other properties, as well as the novel binding molecules themselves.

[0002] The invention further relates to methods applying these molecules in all their versatility.

[0003] In modern biotechnology, one of the most promising and in a number of cases proven applications relies on affinity of proteinaceous molecules for all kinds of substances and/or targets. Proteinaceous binding molecules have been applied in purification of substances from mixtures, in diagnostic assays for a wide array of substances, as well as in the preparation of pharmaceuticals, etc. Typically, naturally occurring proteinaceous molecules, such as immunoglobulins (or other members of the immunoglobulin superfamily) as well as receptors and enzymes have been used. Also peptides derived from such molecules have been used.

[0004] The use of existing (modified) natural molecules of course provides a limited source of properties that evolution has bestowed on these molecules. This is one of the reasons that these molecules have not been applied in all the areas where their use can be envisaged. Also, because evolution always results in a compromise between the different functions of the naturally occurring binding molecules, these molecules are not optimized for their envisaged use.

[0005] Typically, the art has moved in the direction of altering binding properties of existing (modified) binding molecules. In techniques such as phage display of (single chain) antibodies almost any binding specificity can be obtained. However, the binding regions are all presented in the same context. Thus the combination of binding region and its context is often still not optimal, limiting the use of the proteinaceous binding molecules in the art.

[0006] The present invention provides a versatile context for presenting desired affinity regions. The present invention provides a structural context that is designed based on a common structural element (called a core structure) that has been identified herein to occur in numerous binding proteins. This so called common core has now been produced as a novel proteinaceous molecule that can be provided with one or more desired affinity regions.

[0007] This proteinaceous structure does not rely on any amino acid sequence, but only on common structural elements. It can be adapted by providing different amino acid sequences and/or amino acid residues in sequences for the intended application. It can also be adapted to the needs of the particular affinity region to be displayed. The invention thus also provides libraries of both structural contexts and affinity regions to be combined to obtain an optimal proteinaceous binding molecule for a desired purpose.

[0008] Thus the invention provides a synthetic or recombinant proteinaceous molecule comprising a binding peptide and a core, said core comprising a b-barrel comprising at least 4 strands, wherein said b-barrel comprises at least two b-sheets, wherein each of said b-sheet comprises two of said strands and wherein said binding peptide is a peptide connecting two strands in said b-barrel and wherein said binding peptide is outside its natural context. We have identified this core structure in many proteins, ranging from galactosidase to human (and e.g. camel) antibodies with all kinds of molecules in between. Nature has apparently designed this structural element for presenting desired peptide sequences. We have now produced this core in an isolated form, as well as many variants thereof that still have the same or similar structural elements. These novel structures can be used in all applications where other binding molecules are used and even beyond those applications as explained herein. The structure comprising one affinity region (desired peptide sequence) and two b-sheets forming one b-barrel is the most basic form of the invented proteinaceous binding molecules. (proteinaceous means that they are in essence amino acid sequences, but that side chains and/or groups of all kinds may be present; it is of course possible, since the amino acid sequence is of less relevance for the structure to design other molecule of non proteinaceous nature that have the same orientation is space and can present peptidic affintiy regions; the orientation in space is the important parameter). The invention also discloses optimised core structures in which less stable amino acids are replaced by more stable residues (or vice versa) according to the desired purpose. Of course other substitutions or even amino acid sequences completely unrelated to existing structures are included, since, once again, the important parameter is the orientation of the molecule in space. According to the invention it is preferred to apply a more advanced core structure than the basic structure, because both binding properties and structural properties can be designed better and with more predictive value. Thus the invention preferably provides a proteinaceous molecule according the invention wherein said b-barrel comprises at least 5 strands, wherein at least one of said sheets comprises 3 of said strands, more preferably a proteinaceous molecule according to the invention, wherein said b-barrel comprises at least 6 strands, wherein at least two of said sheets comprises 3 of said strands. Variations wherein one sheet comprises only two strands are of course also possible. In an alternative embodiment the invention provides a proteinaceous molecule according to the invention, wherein said b-barrel comprises at least 7 strands, wherein at least one of said sheets comprises 4 of said strands.

[0009] Alternatively the invention provides a proteinaceous molecule according to the invention wherein said b-barrel comprises at least 8 strands, wherein at least one of said sheets comprises 4 of said strands.

[0010] In another embodiment a proteinaceous molecule according to the invention, wherein said b-barrel comprises at least 9 strands, wherein at least one of said sheets comprises 4 of said strands is provided. In the core structure there is a more open side where nature displays affinity regions and a more closed side, where connecting sequences

are present. Preferably at least on affinity region is located at said more open side.

[0011] Thus the invention provides a proteinaceous molecule according to the invention, wherein said binding peptide connects two strands of said b-barrel on the open side of said barrel. Although the location of the desired peptide sequence (affinity region) may be anywhere between two strands, it is preferred that the desired peptide sequence connects the two sheets of the barrel. Thus the invention provides a proteinaceous molecule according to the invention, wherein said binding peptide connects said at least two b-sheets of said barrel. Although one affinity region may suffice it is preferred that more affinity regions are present to arrive at a better binding molecule. Preferably, these regions are arranged such that they can cooperate in binding (e.g. both on the open side of the barrel). Thus the invention provides a proteinaceous molecule according to the invention, which comprises at least one further binding peptide. A successful core element in nature is the one having three affinity regions and three connecting regions. This core in its isolated form is a preferred embodiment of the present invention. However, because of the versatility of the presently invented binding molecules, the connecting sequences on the less open side of the barrel can be used as affinity regions as well. Also other exposed sequences can be affinity regions, for instance sequences exposed to the exterior of the core element. This way a very small bispecific binding molecule is obtained. Thus the invention provides a proteinaceous molecule according the invention, which comprises at least 4 binding peptides.

[0012] Bispecific herein means that the binding molecule has the possibility to bind to two target molecules (the same or different). As already stated it is an object of the present invention to optimise binding molecules both in the binding properties and the structural properties (such as stability under different circumstances (temperature,pH, etc), the antigenicity, etc.). This is done, according tot the invention by taking at least one nucleic acid according to the invention (encoding a proteinaceous binding molecule according to the invention) and mutating either the encoded structural regions or the affinity regions or both and testing whether a molecule with desired binding properties and structural properties has been obtained. Thus the invention provides a method for identifying a proteinaceous molecule with an altered binding property, comprising introducing an alteration in the core of proteinaceous molecules according to the invention, and selecting from said proteinaceous molecules, a proteinaceous molecule with an altered binding property, as well as a method for identifying a proteinaceous molecule with an altered structural property, comprising introducing an alteration in the core of proteinaceous molecules according to the invention, and selecting from said proteinaceous molecules, a proteinaceous molecule with an altered structural property. These alterations can vary in kind, an example being a post-translational modification. The person skilled in the art can design other relevant mutations.

[0013] As explained the mutation would typically be made by mutating the encoding nucleic acid and expressing said nucleic acid in a suitable system, which may be bacterial, eukaryotic or even cell-free. Once selected one can of course use other systems than the selection system.

[0014] The invention also provides methods for producing nucleic acids encoding proteinaceous binding molecules according to the invention, such as a method for producing a nucleic acid encoding a proteinacous molecule capable of displaying at least one desired peptide sequence comprising providing a nucleid acid sequence encoding at least a first and second structural region separated by a nucleic acid sequence encoding said desired peptide sequence or a region where such a sequence can be inserted and mutating said nucleic acid encoding said first and second structural regions to obtain a desired nucleic acid encoding said proteinacous molecule capable of displaying at least one desired peptide sequence and preferably a method for displaying a desired peptide sequence, providing a nucleic acid encoding at least a two b-sheets, said , said b-sheets forming a b-barrel, said nucleic acid comprising a region for inserting a sequence encoding said desired peptide sequence, inserting a nucleic acid sequence comprising a desired peptide sequence, and expressing said nucleid acid whereby said b sheets are obtainable by a method as described above. The invention further provides the application of the novel binding molecules in all fields where binding molecules have been envisaged until today, such as separation of substances from mixtures, typically complex biological mixtures, such as body fluids or secretion fluids, such as blood or milk, or serum or whey.

[0015] Of course pharmaceutical uses and diagnostic uses are clear to the person skilled in the art. In diagnostic uses labels may be attached to the molecules of the invention. In pharmaceutical uses other moieties can be coupled to the molecules of the invention. In both cases this may be chemically or through recombinant fusion. Diagnostic applications and pharmaceutical applications have been described in the art in great detail for conventional binding molecules. For the novel binding molecules according tot the invention no further explanation is necessary for the person skilled in the art. Gene therapy in a targeting format is one of the many examples wherein a binding molecule according to the invention is provided on the gene delivery vehicle (which may be viral (adenovirus, retrovirus, adeno associated virus or lentivirus, etc.) or non viral (liposomes and the like). Genes to be delivered are well known in the art.

[0016] The gene delivery vehicle can also encode a binding molecule according to the invention to be delivered to a target, possibly fused to a toxic moiety. Conjugates of toxic moieties to binding molecules are also well known in the art and are included for the novel binding molecules of the invention.

[0017] The invention will be explained in more detail in the following detailed description.

[0018] The present invention relates to the design, construction, production, screening and use of proteins that contain one or more regions that may be involved in molecular binding. The invention also relates to naturally occurring

proteins provided with artificial binding domains, re-modelled natural occurring proteins provided with extra structural components and provided with one or more artificial binding sites, re-modelled natural occurring proteins disposed of some elements (structural or others) provided with one or more artificial binding sites, artificial proteins containing a standardized core structure motif provided with one or more binding sites. All such proteins are called VAPs (Versatile Affinity Proteins) herein. The invention further relates to novel VAPs identified according to the methods of th invention and the transfer of binding sites on naturally occurring proteins that contain a similar core structure. 3D modelling of such natural occurring proteins can be desired before transfer in order to restore or ensure antigen binding capabilities by the affinity regions present on the selected VAP. Further, the invention relates to processes that use selected VAPs, as described in the invention, for purification, removal, masking, liberation, inhibition, stimulation, capturing, etc,of the chosen ligand capable of being bound by the selected VAP(s).

LIGAND BINDING PROTEINS

[0019] Many naturally occurring proteins that contain a (putative) molecular binding site comprise two functionally different regions: The actual displayed binding region and the region(s) that is (are) wrapped around the molecular binding site or pocket, called the Core Structure (CS) also called scaffold herein. These two regions are different in function, structure, composition and physical properties. The CS structures ensures a stable 3 dimensional conformation for the whole protein, and act as a steppingstone for the actual recognition region. Two functional different classes of ligand binding proteins can be discriminated. This discrimination is based upon the presence of a genetically variable or invariable ligand binding region. In general, the invariable ligand binding proteins contain a fixed number, a fixed composition and an invariable sequence of amino acids in the binding pocket in a cell of that species. Examples of such proteins are all cell adhesion molecules, e.g. N-CAM and V-CAM, the enzyme families, e.g. kinases and proteases and the family of growth receptors,e.g EGF-R, bFGF-R. In contrast, the genetically variable class of ligand binding proteins is under control of an active genetic shuffling-, mutational or rearrangement mechanism enabling an organism or cell to change the number, composition and sequence of amino acids in, and possibly around, the binding pocket. Exmples of these are all types of light and heavy chain of antibodies, B-cell receptor light and heavy chains and T-cell receptor alfa, beta, gamma and delta chains. The molecular constitution of wild type CS's can vary to a large extent. For example, Zinc finger containing DNA binding molecules contain a totally different CS (looking at the amino acid composition) than antibodies although both proteins are able to bind to a specific target.

SCAFFOLDS (CORE STRUCTURES) AND LIGAND BINDING DOMAINS

Antibodies obtained via immunizations

[0020] The class of ligand binding proteins that express variable (putative) antigen binding domains has been shown to be of great value in the search for ligand binding proteins. The classical approach to generate ligand binding proteins makes use of the animal immune system. This system is involved in the protection of an organism against foreign substances. One way of recognizing, binding and clearing the organism of such foreign highly diverse substances is the generation of antibodies against these molecules. The immune system is able to select and multiply antibody producing cells that recognize an antigen. This process can also be mimicked by means of active immunizations. After a series of immunizations antibodies may be formed that recognize and bind the antigen. The possible number of antibodies that can be formed due to genetic rearrangements and mutations, exceeds the number of 1040. However, in practice, a smaller number of antibody types will be screened and optimized by the immune system. The isolation of the correct antibody producing cells and subsequent immortalization of these cells or, alternatively, cloning of the selected antibody genes directly, antigen-antibody pairs can be conserved for future (commercial and non-commercial) use.

[0021] The use of antibodies obtained this way is restricted only to a limited number of applications. The structure of animal antibodies is different than antibodies found in human. The introduction of animal derived antibodies in humans will almost certainly cause immune responses adversing the effect of the introduced antibody (e.g. HAMA reaction). As it is not allowed to actively immunize men for commercial purposes, it is not or only rarely possible to obtain human antibodies this way. Because of these disadvantages methods have been developed to bypass the generation of animal specific antibodies. One example is the removal of the mouse immune system and the introduction of the human immune system in such mouse. All antibodies produced after immunization are of human origin. However the use of animal has also a couple of important disadvantages. First, animal care has a growing attention from ethologists, investigators, public opinion and government. Immunization belongs to a painful and stressful operation and must be prevented as much as possible. Second, immunizations do not always produce antibodies or do not always produce antibodies that contain required features such as binding strength, antigen specificity, etc. The reason for this can be multiple: the immune system missed by co-incidence such a putative antibody; the initially formed antibody appeared

to be toxic or harmful; the initially formed antibody also recognizes animal specific molecules and consequently the cells that produce such antibodies will be destroyed; or the epitope cannot be mapped by the immune system (this can have several reasons).

Otherwise obtained antibodies

[0022] It is clear, as discussed above, that immunization procedures may result in the formation of ligand binding proteins but their use is limited, inflexible and uncontrollable. The invention of methods for the bacterial production of antibody fragments (Skerra and Pluckthun, 1988; Better et al., 1988) provided new powerful tools to circumvent the use of animals and immunization procedures. It is has been shown that cloned antibody fragments, (frameworks, affinity regions and combinations of these) can be expressed in artificial systems, enabling the modulation and production of antibodies and derivatives (Fab, VL, VH, scFv and VHH) that recognize a (putative) specific target in vitro. New efficient selection technologies and improved degeneration strategies directed the development of huge artificial (among which human) antibody fragment libraries. Such libraries potentially contain antibodies fragments that can bind one or more ligands of choice. These putative ligand specific antibodies can be retrieved by screening and selection procedures. Thus, ligand binding proteins of specific targets can be engineered and retrieved without the use of animal immunizations.

Other immunoglobulin superfamily derived scaffolds

[0023] Although most energy and effort is put in the development and optimization of natural derived or copied human antibody derived libraries, other scaffolds have also been described as succesful scaffolds as carriers for one or more ligand binding domains. Examples of scaffolds based on natural occurring antibodies encompass minibodies (Pessi et al., 1993), Camelidae VHH proteins (Davies and Riechmann, 1994; Hamers-Casterman et al., 1993) and soluble VH variants (Dimasi et al., 1997; Lauwereys et al., 1998). Two other natural occurring proteins that have been used for affinity region insertions are also member of the immunoglobulin superfamily: the T-cell receptor chains (Holler et al., 2000) and fibronectin domain-3 regions (Koide et al., 1998). The two T-cell receptor chains can hold three affinity regions each while for the fibronectin region the investigators described only two regions. Non-immunoglobulin derived scaffolds Besides immunoglobulin domain derived scaffolds, non-immunoglobulin domain containing scaffolds have been investigated. All proteins investigated contain only one protein chain and one to four affinity related regions. Smith and his colleagues (1998) reported the use of knottins (a group of small disulfide bonded proteins) as a scaffold. They successfully created a library based on knottins that had 7 mutational amino acids. Although the stability and length of the proteins are excellent, the low number of amino acids that can be randomized and the singularity of the affinity region make knottin proteins not very powerful. Ku and Schultz (1995) successfully introduced two randomized regions in the four-helix-bundle structure of cytochrome b562. However, selected binders were shown to bind with micromolar Kd values instead of the required nanomolar or even better range. Another alternate framework that has been used belongs to the tendamistat family of proteins. McConnell and Hoess (1995) demonstrated that alpha-amylase inhibitor (74 amino acid beta-sheet protein) from Streptomyces tendae could serve as a scaffold for ligand binding libraries. Two domains were shown to accept degenerated regions and function in ligand binding. The size and properties of the binders showed that tendamistats could function very well as ligand mimickers, called mimetopes. This option has now been exploited. Lipocalin proteins have also been shown to be succesful scaffolds for a maximum of four affinity regions (Beste et al., 1999; Skerra, 2000 BBA; Skerra, 2001 RMB). Lipocalins are involved in the binding of small molecules like retinoids, arachidonic acid and several different steroids. Each lipocalin has a specialized region that recognizes and binds one or more specific ligands. Skerra (2001) used the lipocalin RBP and BBP to introduce variable regions at the site of the ligand binding domain. After the construction of a library and successive screening, the investigators were able to isolate and characterize several unique binders with nanomolar specificity for the chosen ligands. It is currently not known how effective lipocalins can be produced in bacteria or fungal cells. The size of lipocalins (about 170 amino acids) is pretty large in relation to VHH chains (about 100 amino acids) which might be too large for industrial applications.

[0024] The VAPs of the invention can be used in an enormous variety of applications, from therapeutics to antibiotics, from detection reagents to purification modules, etc. etc. In each application the environment and the downstream applications determines the features that a ligand binding protein should have, e.g. temperature stability, protease resistance, tags, etc. What ever the choice of the scaffolds is, all have their own unique properties. Some properties can be advantageous for certain applications while others are unacceptable. For large scale industrial commercial uses it is crucial that scaffolds contain a modular design in order to be able to mutate, remove, insert and swap regions easily and quick. Modularity make it possible to optimize for required properties via standardized procedures and it allows domain exchange programs, e.g. exchange of pre-made cassettes. As optimal modular scaffold genes should meet certain features, they have to be designed and synthetically constructed while it is very unlikely that natural

retrieved genes contains such features.

[0025] Besides modularity there are several other properties that should be present or just absent in the scaffold gene or protein. All scaffold systems that are based on frameworks that are present in natural proteins inherit also their natural properties. These properties have been optimized by evolutionary forces for the system in which this specific protein acts. Specific properties encompass for example codon usage, codon frequency, expression levels, folding patterns and cystein bridge formation. Industrial commercial production of such proteins, however, demands optimal expression, translation and folding to achieve economic profits. Not only should the genetic information be compatible and acceptable for the production organism, protein properties should also be optimal for the type of application. Such properties can be heat sensitivity, pH sensitivity, salt concentration sensitivity, proteolytic sensitivity, stability, purification possibilities, and many others.

[0026] Thus, to be of practical use in affinity processes, specific binding activity alone is not sufficient. The specific binding agent must also be capable of being linked to a solid phase such as a carrier material in a column, an insoluble bead, a plastic, metal or paper surface or any other useful surface. Ideally, this linkage is achievable without any adverse effects on the specific binding activity. Therefore the linkage is preferably accomplished with regions in the VAP molecule that are relatively remote from the specific affinity regions.

[0027] An important embodiment of the invention is an affinity-absorbant material comprising a specific binding agent immobilised on a porous silica or the like, the specific binding agent comprising a selection of VAP molecules.

A particularly important embodiment of the invention is an affinity-absorbant material comprising a special binding agent immobilised on a porous carrier material, such as silica or an inert, rigid polymer or the like, having a pore size of at least 30A but not greater than 1000A, wherein the specific binding agent comprises a selection of VAP molecules. Preferably, the carrier has a pore size of at least 60A. Preferably, the pore size is not greater than 500A, and more preferably, not greater than 300A.

[0028] The pore size of a carrier medium such as silica or inert polymers can be determined using e.g. standard size exclusion techniques or other published methods. The nominal pore size is often referred to as the mean pore diameter, and is expressed as a function of pore volume and surface area, as calculated by the Wheeler equation (MPD= (40,000 x pore volume)/surface area. The pore volume and surface area can be determined by standard nitrogen absorption methods of Brunauer, Emmett and Teller.

[0029] Products in which VAPs can be applied in a way that leaves the VAPs present up to, and also including, the end product, have examples from a very wide range of products. But also in processes where the VAPs are immobilized and preferably can be regenerated for recycled use, the major advantage of VAPs is fully exploited, i.e. the relative low cost of VAPs that makes them especially suitable for large scale applications, for which large quantities of the affinity bodies need to be used. The list below is given to indicate the scope of applications and is in no way limiting. Product or process examples with possible applications in brackets are;

(1) industrial (food) processing such as the processing of whey, tomato pomace, citrus fruits, etc. or processes related to bulk raw materials of agricultural origin such as the extraction of starch, oil and fats, proteins, fibers, sugars etc. from bulk crops such as, but not limited to; potato, corn, rice, wheat, soybean, cotton, sunflower, sugarbeet, sugarcane, tapioca, rape. Other examples of large process streams are found in the diary-related industries e.g. during cheese and butter manufacturing. As the VAPs can be used in line with existing processing steps and the VAPs do not end up in the final product as a result of their irreversible immobilisation to support-materials, they are exceptionally suited for the large scale industrial environments that are customary in agro-foodprocessing industries.

In a more detailed example, the whey fraction that is the result of the cheese manufacturing processes contain a relatively large number of low-abundant proteins that have important biological functions, e.g. during the development of neonates, as natural antibiotics, food-additives etc. Examples of such proteins are lactoferrin, lactoperoxidase, lysozyme, angiogenine, insulin-like growth factors (IGF), insulin receptor, IGF-binding proteins, transforming growth factors (TGF), bound- and soluble TGF-receptors, epidermal growth factor (EGF), EGF-receptor ligands, interleukine-1 receptor antagonist. Another subclass of valuable compounds that can be recovered from whey are the immunoregulatory peptides that are present in milk or colostrum. Also specific VAPs can be selected for the recovery of hormones from whey. Examples of hormones that are present in milk are; prolactin, somatostatin, oxytocin, luteinizing hormone-releasing hormone, thyroid-stimulating hormone, thyroxine, calcitonin, estrogen, progesterone. Other examples where the VAPs are exceptionally suitable, are large scale affinity chromatography applications where purification of a target compound is limited for economic reasons; the cost of purification becomes too high in relation to the price of the purified compound. A very significant cost reduction can often be achieved through a reduction of the number of purification steps by a far more selective purification step that is based on specific affinities such as displayed by VAPs. In some examples, single step purifications are even feasible, where the target compound is directly purified to acceptable levels from the originating process stream (e.g. single step enzyme purification from extracellular culture media or cell extracts of industrial production micro-

organisms or other cell-based production methods).
(2) edible consumer products such as ice-cream, oil-based products such as oils, margarines, dressings and mayonaisse, other processed foodproducts as soups, sauces, pre-fabricated meals, soft-drinks, beer, win , etc. (preservation and prevention of spoilage, through direct antibiotic activity or selective inhibition of enzymes, protecting sensitive motives during processing, e.g. from enzymes or compounds that influence quality of end products through its presence in an active form, or for enzymes that preferrably needs to be active down-stream of a denaturing process step and where the binding with a specific VAPs would prevent the active site of the enzyme to be denatured) controlled release of flavours and odours, molecular mimics to mask or enhance flavours and odours e.g. masking or removing bitter components in beer brewing industries, removal of pesticides or other contaminants.
(3) personal care products such as shampoos, hair-dying liquids, washing liquids, laundry detergents, gels as applied in different forms such as powders, paste, tablet or liquid form etc. (anti-microbial activity for inhibition of dandruff or other skin-related microbes, anti-microbial activity for toothpastes and mouthwashes, increased specificity for stain-removing enzymes in detergents, stabilizing labile enzymes in soaps or detergents to increase e. g. temperature or pH stability, increased binding activity for hair-dye products, inhibiting enzymes that cause body odours, either in skin applications or in clothing accessories such as shoe-inlays, hygiene tissues)
(4) non-food applications such as printing inks, glues, paints, paper, hygiene tissues etc. (surface-specific inks, glues, paints etcetera for surfaces that are otherwise difficult to print e.g. most current plastics, but especially those that are based on polyofines bottles or containers, or for surfaces where highly specific binding is required, e.g. lithographic processes in electronic chip manufacturing, authentication of value papers, etc)
(5) environmental protection processes such as water purification, bioremediation, clean-up of process waters, concentration of contaminants (removal of microorganisms, viruses, organic pollutants in water purification plants or e.g. green-house water recycling systems, removal of biological hazards from air-ventilation ducts)
(6) animal feed products in dry or wet forms (removal, masking or otherwise inhibiting the effects of anti-nutritional factors that often occur in feed components both for cattle and fish farming, notably protease inhibitors or negative factors such as phytic acid, addition of VAPs as antimicrobial agents to replace current antibiotics with protein-based antibiotics) Although the preferred embodiments of this patent include industrial processes, the use of VAPs in a manner of affinity chromatography is certainly not limited to these applications. On the "low volume / high value" side of the scale, a variety of applications is feasible for pharmaceutical, diagnostic and research purposes where price is of lesser importance for application, but the availability of VAPs against ligands that are notoriously difficult to raise antibodies against in classical immune responses. Also the small size and high stability will provide "low volume / high value" applications were VAPs are superior to conventional antibodies or fragments thereof.
(7) pharmaceutical aplications where VAPs can be used as therapeutics themselves, particularly when the core is designed to resemble a natural occurring protein, or to identify and design proper affinity regions and/or core regions for therapeutics.
(8) diagnostic applications where VAPs, as a result of their 3D structure that differ in essential ways from commonly used antibodies or antibody fragments, may detect a different class of molecules. Examples are the detection of infectious prions, where the mutation causing the infectious state is buried inside the native molecule. Coventional antibodies can only discriminate the infectious form under denatured conditions, while the small and exposed AR's of VAPs are able to recognize more inwardly placed peptide sequences.
(9) research applications where VAPs are bound to e.g. plate surfaces or tissues to increase detection levels, localize specific compounds on a fixed surface, fix tracer molecules in position etc. or where selected genes that code for specific VAPs are either transiently, continuously or in a controlled manner expressed by translating said genes in a cellular environment, and where through its targeted expression functional knock-outs of target molecules are formed. For example mimicking a receptor ligand may interfere with normal signal-transduction pathways, or VAPs that function as enzyme inhibitors may interfere with metabolic pathways or metabolic routing.

[0030] Diverse as the above examples are, there are commonalities in the ways that VAPs can be applied, as is illustrated by the following categories that form a matrix in combination with the applications:

1. affinity chromatography where VAPs are immobilized on an appropriate support e.g. in chromatography columns that can be used in line, in series or in carroussel configurations for fully continuous operation. Also pipes, tubes, in line filters etc. can be lined with immobilized VAPs. The support material on which the VAPs can be immobilized can be chosen to fit the process requirements in terms of compression stability, flow characteristics, chemical inertness, temperature-, pH- and solvent stability etc. Relatively incompressible carriers are preferred, especially silica or rigid inert polymers. These have important advantages for use in industrial-scale affinity chromatography because they can be packed in columns operable at substantially higher pressures than can be applied to softer carrier materials such as agarose. Coupling procedures for binding proteins to such diverse support materials are

well-known. After charging the column with the process stream of choice, the bound ligands can be desorbed from the immobilized VAPs through well-known procedures such as changes in pH or salt concentrations after which the VAPs can be regenerated for a new cycle. The high stability of selected VAPs makes them exceptionally suitable for such repeated cycles, thus improving the cost efficiency of such recovery and purification procedures. Th principles and versatility of affinity chromatography have been widely described in thousands of different applications.

2. insoluble beads are a different form of affinity chromatography where the support material on which VAPs are immobilized are not fixed in position but are available as beads from for example , silica, metal, magnetic particles, resins and the like. can be mixed in process streams to bind specific ligands in e.g. fluidised beds or stirred tanks, after which the beads can be seperated from the process stream in simple procedures using gravity, magnetism, filters etc.

3. coagulation of target ligands by crosslinking the ligands with VAPs, thereby reducing their solubility and concentrating the ligands through precipitation. For this purpose, VAPs should be bivalent, i.e. at least two AR's must be constructed on either side of the scaffold. The two AR's can have the same molecular target but wo different molecular targets are preferred to provide the cross-linking or coaggulation effects.

4. masking of specific molecules to protect sensitive motives during processing steps, to increase the stability of the target ligand for adverse pH, temperature or solvent conditions, or to increase the resistance against deteriorating or degrading enzymes. Other functional effects of molecular masking can be the masking of volatile molecules to alter the sensory perception of such molecules. In contrast the slow and conditional release of such molecules from VAPs can be invisaged in more down-stream processing steps, during consumption or digestion or after targeting the VAPs-ligand complex to appropriate sites for biomedical or research applications. Also molecular mimics of volatile compounds using VAPs with specific receptor binding capacity can be used to mask odours from consumer products.

5. coating of insoluble materials with VAPs to provide highly specific surface affinity properties or to bind VAPs or potential fusion products (i.e. products that are chemically bound to the VAPs or, in case of protein, are co-translated along with the VAPs in such manner that the specicifity of the VAPs remains unchanged) to specific surfaces. Examples are the use of VAPs to immobilize specific molecules to e.g. tissues, on plates etc. to increase detection levels, localize specific compounds on a fixed surface, fix tracer molecules in position etc.

[0031] Certainly not all natural CS's are interesting from a commercial and/or industrial point of view. For example, the stability and sensitivity of the whole protein should meet the requirements that go along with the proposed application. Ligand bindig proteins in an acidic environment are not per se useful in high salt or high temperature environments. It is not possible to design one scaffold that has all possible features to function as a one for all scaffold. For example, there are applications that require proteolytic insensitive scaffolds while other applications require specific protease cleavage sites in the scaffold. For these and many other applications it is not possible to design one scaffold that meets all requirements. However, most differences in scaffold function have to do with properties of the amino acid that are exposed on the outside of the protein. For example pH effect, proteolytic recognition and cleavage, solubility, expression levels, antigenicity and many more features mostly have to do with amino acid side chain properties instead of the core of the scaffold and the backbone. Changing the properties of these amino acids will typically have a minor effect on the overall structure of a scaffold. Thus is possible to change application forms by changing the quality of amino acids located on the outside of the scaffold or of those amino acids that are not involved in ligand binding or structure determination. With this in mind it is possible to design and construct scaffolds that can be used in multiple kinds of ligand binding environments without changing the properties and spatial position of the ligand binding domain. Consequence of this idea is that affinity domains can be swapped from one scaffold to another scaffold, if the scaffolds contain similar CS's, without losing their specificity for the ligand. We call this swapping technology Modular Affinity and Scaffold Transfer technology (MAST technology).

Adapting scaffolds and affinity regions through mutations.

[0032] One method to introduce random mutations in pieces of DNA is the introduction of inosine residues in DNA fragments. Inosine is widely used as a residue in sequencing methods and degenerated primers. Inosine is a form of deoxy-nucleotide-tri-phosphate that can be incorporated in DNA fragments without disturbing the overall DNA structure but it lacks side groups that are responsible for H-H bridge formation. Typically, adenosine can form 2 H-H bridges with thymidine while cytosine can form 3 H-H bridges with guanidine their pairs are opposite positioned. Therefore inosine can be found opposite to A, C, G or T. The effect of inosine in DNA strands is that it will not contribute to the melting temperature of the DNA fragment. If inosine is introduced in DNA strands cellular DNA repair mechanisms will try to correct this and replace the inosine will the corresponding complementary required nucleotide of the opposite DNA strand. However under in vitro conditions, environmental systems determine the acceptance, incorporation or removal

of inosine residues. For example, proof reading DNA polymerases are able to correct DNA strands for inosine residues but non-proof reading polymerases lack this property.

In DNA amplification reactions inosine residues can be introduced by at random during the formation of a new strand of DNA. Most non-proof reading polymerases (even thermostable ones) can also, by chance, incorporate inosine residues, instead of one of the four regular dNTPs, at random positions during the second strand assembly. During further steps a random regular nucleotide can be placed at the base site opposite to the inosine residue. The number of mutations is Gaussian distributed. Factors that can influence the relative number of mutations are PCR cycli number, concentration dITP, concentration regular dNTPs, polymerase properties, DNA concentration and PCR buffer composition. An optimized procedure can be applied to every stretch of nucleotides flanked by one or more primer regions. Inosine based mutational strategies have been reported for a couple of target sequences. The inosine procedure has a couple of explicit important features: it does not introduce frameshifts, it keeps the original length of the DNA fragments, it is unbiased for the mutational position, it is (relatively) unbiased for the mutational type (dNTP type), easy to perform, controllable and it is applicable to every part of DNA that can be amplified. The effects of the mutated DNA can be: change in DNA properties (e.g. promoter elements), RNA stability and most important for our purposes, changes in amino acid composition and sequence of the DNA fragment. By influencing the number of mutations and the domains of mutations, new protein domain properties can be generated. By controlling the relative number of changes it is possible to keep the overall amino acid composition. This way structural important amino acids or stretches of amino acids within the corresponding mutated domain can be kept for a certain percentage of the obtained clones. However, the random introduction of inosine can also by coincidence result in the introduction of undesired stop codons. By controlling PCR conditions it is possible to restrict the average number of stop codons. In addition, selection procedures (frame and stop codon dependent) and the use of specific strains that can overrule specific stop codons, e.g. amber or ochre mutations in specific E.coli strains, can at least partly compensate for the effect of the putative introduced stop codons.

The consequence that only a limited number of amino acids will changed creates a new approach in the generation of randomization in for example affinity regions (e.g. CDR3). Now structural information and/or critical residues or properties at certain positions, in at least a significant percentage of clones formed using dITP mutations, can be conserved and can be extremely important in building high quality antibody libraries.

The combination of a method for producing in frame mutations within certain pre-determined regions using PCR or other techniques that include inosine residues and the possibility to (functionally) screen large numbers of mutants make an excellent and powerful combination not only in optimizing affinity regions in antibodies but also in other affinity regions or even in the optimization in the functionality or specificity of proteinous domains. In addition the combination of these techniques can be used to optimize functional domains in stretches of DNA or RNA. For example promoter, UTR, leaders, structural units, transcription initiation and transcriptional termination, translational effects, etc.

The size of the region that can be affected by the inosine mutational method is not restricted as long as the DNA fragment can be amplified.

[0033] In a specific example, it is possible now to generate structures, based on a template, that have new improved or adapted properties that were not present in the original template. Optimal temperture sensitivity, optimal protease resistance, optimal ligand binding, optimal ligand specificity, optimal pH sensitivity, optimal solubility, optimal expression, optimal translation, optimal secretion, optimal codon usage, optimal functionality, etc in which the term optimal represents the optimal configuration for a certain application (both positive and negative).

CORE STRUCTURE DEVELOPMENT

[0034] In commercial industrial applications it is very interesting to use single chain peptides, instead of multiple chain peptides because of low costs and high efficiency of such peptides in production processes. One example that could be used in industrial applications are the VHH antibodies. Such antibodies are very stable, can have high specificities and are relatively small. However, the scaffold has evolutionarily been optimized for an immune dependent function but not for industrial applications. In addition, the highly diverse pool of framework regions that are present in one pool of antibodies prevents the use of modular optimization methods. Therefore a new scaffold was designed based on the favourable stability of VHH proteins.

3D-modelling and comparative modelling software was used to design a scaffold that meets the requirements of versatile affinity proteins (VAPs). Because it is not possible yet to predict protein structures, protein stability and other features, it is necessary to test the computer designed and modelled scaffolds. Multiple primary scaffolds were constructed and pooled. All computer designed proteins are just an estimated guess. One mutation or multiple amino acid changes in the primary scaffold may make it a successful scaffold or make it function even better than predicted. To accomplish this the constructed primary scaffolds are subjected to a mild mutation program.

[0035] Hereto the inosine dependent mutational technology is applied. Inosine insertions at random positions can ultimately change the amino acid compositions and amino acid sequences of the primary scaffolds. This way new

(secondary) scaffolds are generated. In order to test the functionality, stability and other required or desired features of the scaffolds, a set of known affinity regions are inserted in the secondary scaffolds. All functional scaffolds that display the inserted affinity regions in a correct fashion will be able to bind to the ligand that corresponds to the inserted affinity regions. Phage diplay techniques help to conserve the genetic and phenotypic information of the scaffolds. After panning procedures optimal and functional scaffolds are retrieved. These scaffolds can be used for the insertion of new affinity regions.

INITIAL AFFINITY REGIONS FOR LIBRARY CONSTRUCTION

[0036] In the present invention new and unique affinity regions are required. Affinity regions can be obtained from natural sources, degenerated primers or stacked DNA triplets. All of these sources have certain important limitations as described above. In our new setting we designed a new and strongly improved source of affinity regions which have less restrictions, can be used in modular systems, are extremely flexible in use and optimization, are fast and easy to generate and modulate, have a low percentage of stop codons, have an extremely low percentage of frameshifts and wherein important structural features will be conserved in a large fraction of the new formed clones and new structural elements can be introduced.

[0037] The major important affinity region (CDR3) in both light and heavy chain in normal antibodies has a average length between 11 (mouse) and 13 (human) amino acids. Because in such antibodies the CDR3 in light and heavy chain cooperatively function as antigen binder, the strength of such a binding is a result of both regions together. In contrast, the binding of antigens by VHH antibodies (Camelidae) is a result of one CDR3 regions due to the absence of a light chain. With an estimated average length of 16 amino acids these CDR3 regions are significantly longer than regular CDR3 regions (Mol. Immunol. Bang Vu et al., 1997, 34, 1121-1131). It can be emphasized that long or multiple CDR3 region have potentially more interaction sites with the ligand and can therefore be more specific and bind with more strength. Average lengths of the major affinity region(s) should preferably be about 16 amino acids. In order to cover as much as possible potentially functional CDR lengths the major affinity region can vary between 1 and 50 or even more amino acids. As the structure and the structural classes of CDR3 regions (like for CDR1 and CDR2) have not been clarified and understood it is not possible to design long affinity regions in a way that the position and properties of crucial amino acids are correct. Therefore, most libraries were supplied with completely degenerated region in order to find at least some correct regions. In the invention we describe the use of natural occurring VHH CDR3 region as a template for new affinity regions. CDR3 regions were isolated from mRNA coding for VHH antibodies originating to various animals of the camelidae group by means of PCR techniques. Next this pool of about 108 different CDR3 regions, which differ in the coding for amino acid composition, amino acid sequence, putative structural classes and length, is subjected to a mutational process by PCR amplification that includes dITP in the reaction mixture. The result is that the very most products will differ from the original templates and thus contain coding regions that potentially have different affinity region. Other very important consequences are that the products keep their length, the pool keeps their length distribution, a significant part will keep structural important information while others might form non-natural classes of structures, the products do not or only rarely contain frame shifts and a the majority of the products will lack stop codons. These new affinity regions can be cloned into the selected scaffolds. The newly constructed library can be subjected to screening procedures similarly as regular libraries known by an experienced user in the field of the art.

AFFINITY MATURATION

[0038] After one or more selection rounds, an enriched population of VAPs is formed that recognizesthe ligand selected for. In order to obtain better, different or otherwise changed VAPs against the ligand(s), the VAP coding regions or parts thereof can be the subject of a mutational program by several means. Beside chain shuffling, error prone PCR stratagies or mutator strains, the inosine dependent mutational technology can be applied. Due to the modular nature of the VAPs, parts of the VAP can be used in the last mentioned mutational technology. Several stratagies are possible: First, the whole VAP or VAPs can be used as a template. Second, only one or more affinity region can be mutated. Third framework regions can be mutated. Fourth, fragments throughout the VAP can be used as a template. Of course itterative processes can be applied to change more regions. The average number of mutations can be varied by changing PCR conditions. This way every desired region can be mutated and every desired level of mutation numbers can be applied independently. After the mutational procedure, the new formed pool of VAPs can be re-screened and re-selected in order to find new and improved VAPs against the ligand(s). The process of maturation can be re-started and re-applied as much rounds as necessary.

The effect of inosine dependent mutational is that not only affinity regions 1 and 2 with desired affinities and specificities can be found but also that minor changes in the selected affinity region 3 can be inttroduced. It has been shown that mutational programs in this major ligand binding region can strongly increase ligand binding properties. In conclusion, the invention described here is extremely powerful in the maturation phase.

EXAMPLES

Example 1.

INTRODUCING NEW BINDING AND/OR OTHER PROPERTIES OF OBTAINED CAMEL ANTIBODIES.

[0039] Lama pacos lymphocytes are obtained by isolating lymphocytes via a discontinue gradient of Ficol (Pharmacia Biotech). Five ml heparin lama blood is diluted with 5 ml phosphate buffered saline (PBS) and applied on the top of 4 ml Ficol. After centrifugation for 20 minutes at 1000g without brakes an interphase will form mainly containing lyphocytes. This fraction is isolated and diluted with PBS to 50 ml and the cell are pelleted at 1000g by centrifugation for 10 minutes. The supernatant is discarded and the cells are lysed in Tri-Pure (Roche Diagnostics) using a teflon piston and the mRNA is isolated according to the manufactures protocol. Two microgram mRNA is reverse transcribed using oligo-dT primers and AMV reverse transcriptase (Roche-Diagnostics) according to the manufactures procedure. Five microliter of the obtained cDNA is directly used in a PCR amplification reaction for 30 cycles at 93°C 15 seconds, 50°C 20 seconds, 70°C 60 seconds with a mixture containing 250nM forward and backward primers (ordered at Genset), 100micromolar PCR grade dNTPs (Roche-Diagnostics), 2.5 units High Fidelity Taq mix (Roche-Diagnostics) and the supplied buffer including magnesium sulphate. The produced products are separated on a 1% agarose Tris Buffered EDTA (TBE) gel and cut out and purified using Qiagen PCR purification methods. The isolated PCR products are used in mutational PCR reactions. The cycles are repeated for 10 times using the same temperature and time conditions as described above in the RT-PCR reaction. The mixtures contain dITP concentrations varying from 1 micromolar to 200 micromolar, 10 micromolar dNTPs, 2.5 unit Taq (Roche-Diagnostics), 5' biotin labelled primers at 250nM each (same primers sequences as mentioned above). The formed PCR products are purified using a biotin purification kit according to the manufactures procedures. Inosines are replaced by adding short random phosphorylated oligomers to the template. Mixture is heated and cooled down to room temperture. After the addition of buffer, dNTPs, DNA T4 ligase and the klenow fragment DNA polymerase and an incubation at room temperature a proof reading polymerase is added to replace the inosines of the template DNA that is now paired with a non-inosine containing duplicate build from the random oligos and filled in with the polymerase. Isolated products are re-PCRed for 10 cycles according to the first PCR schema with the exception that the primers now contain an unique restriction site at the 5 prime end to enable unidirectional cloning. After amplification the biotin labelled template used in this PCR reaction is removed using the above mentioned method. Amplified products are purified using Qiagen PCR purification kit and digested with appropriate restriction enzymes. Next the gel separated, cut out and repurified DNA fragments are ligated in a bla-containing (ampicillin resistance gene) phage display vector expression vector with the M13 g3 protein in frame along with a His tag, a detection tag (HA or VSV) and a proteolytic site (trypsin) which are located in between the PCR cloning site and the g3 open reading frame. After electroporation of the ligation products in TG1 E. coli, cells are grown on plates that contain 4% glucose, ampicillin and 2*TY-agar. Harvested cells are used to infect with the kanamycine bearing VCSM13 helper phage (Stratagene). Helper phage infected cells are grown in 2*TY/ kanamycine/ ampicillin on a 200rpm shaking platform at 30°C for 8 hours. Next the bacteria are removed by pelleting at 20000g at 4°C for 30 minutes. The supernatant is filtered through a 0.45 micrometer PVDF filter membrane (Milli-Pore). Poly-ethylene-glycol and NaCl are added to a final concentration of the flow through to respectively 4% and 0.5M. This way phages will precipitate on ice and can be obtained by centrifugation. The phage pellet is solved in 50% glycerol/50% PBS and stored at -20°C. Panning is performed by diluting about 10^{13} stored phages in 10ml PBS/0.5% gelatine. The antigen is coated on 2 types of immunotubes, low charged and high charged tubes (Greiner). After a blocking period of 2-8 hours at 4°C the phage mixture is added and incubated on a slowly rotating platform at 4°C for 60 minutes. The immunotubes are washed 5 times with PBS/0.5% gelatine and 5 times with PBS/0.05% tween-20. Bound phages are eluted with 0.1M glycine pH 2.2 (HCl) for 1 minute and the eluted fluid is neutralized with Tris 8.8. Ten volumes of XL1-blue cells at OD 0.1-0.3 are infected with the neutralized phage solution. After infection the cells are plated on 2*TY agar/ ampicillin/tetracyclin/4%glucose plates and grown overnight at 37°C. Re-panning can be performed until a desired number of cycles or colonies or individual or pooled colonies can be screened for inserts and/or binding capacity and antigen specificity. Growth of individual or pooled clones and the formation of chimeric phages is basically the same as growing a whole bank of phages. After phage precipitation the phage are added to coated and blocked ELISA plates in order to bind to the target antigen followed by extensive washing steps similarly as described above. Anti-M13-HRP antibodies (Pharmacia Biotech) are added (1:5000) to test the number of bound phages (in relation to control wells in the ELISA plate). After a 30 minute binding period and a tenfold wash procedure with PBS/0.05% tween-20, the HRP-coated antibodies against M13 phages are detected using TMB or other ELISA-reader and HRP compatible reagents. The binding properties are read in means of staining and staining intensity. Strong signals indicate strong binding. Selected clones can be re-mutated using the above described dITP method or other mutational strategies in order to improve binding and/or other properties.

Example 2.

DESIGN OF A MAXIMAL PRIMARY SCAFFOLD

[0040] A primary scaffold is designed as a template to generate secondary scaffolds with improved properties or features by mutational strategies. A primary scaffold design has to meet some desired criteria: first it should be highly soluble, very stable and easy to fold. The solubility of proteins can be changed for example by changing amino acids that have side chains that are exposed to the environment. Hydrophilic amino acids like lysine, aspartic acid, arginine, histidine, serine, threonine or glutamic acid can improve the solubility. The stability and folding properties can be improved by the addition of intramolecular bonds. The most powerful bond is found in sulphide-bridges between two opposed cysteine residues. However, bacterial cells, in contrast to prokaryotic cells like yeast cells, can have problems with cysteine bridge formations and therefore only a limited number of cysteine bridges should be incorporated in proteins that have to be expressed in bacterial cells. Because the used scaffold is structurally based upon immunoglobular domains and because at least some bacterial proteins contain such structures or deviations of these, one can incorporate the structural and amino acid sequence knowledge of these proteins into new scaffolds that are based upon these structures. Using Modeller 6.0, Insight II individual or stretches of amino acid changes were checked for stability prognoses in comparative modelling situations. Especially the outer amino acids of the structures were initially modelled. Vector NTI suite 7.0 software was used to design the corresponding DNA sequences including codon optimizations for E.coli cells and the incorporation of desired restriction sites. Variable regions from human mouse and lama (especially VHH) have been used as a template to design a new scaffold. This scaffold therefore contains 9 beta-elements (see figure 1 and 2) arranged in 2 beta-sheets. The four even numbered loops can form a binding region on one side of the scaffold (for example as found in antibodies) while the odd numbered loops on the other side of the scaffold can be used for a binding region for another or even the same molecule. From all models that were created using comparative modelling software the next primary scaffold with 9 beta-elements was chosen based on energy minimization, potential folding properties, potential solubility and potential stability:

N V K L V E K G G N F V E N D D D L K L

AATGTGAAACTGGTTGAAAAAGGTGGCAATTTCGTCGAAAACGATGACGATCTTAAGCTC

T C R A E G * * * * * Y C M G W F R Q A

ACGTGCCGTGCTGAAGGTNNNNNNNNNNNNNNNNTACTGCATGGGTTGGTTCCGTCAGGCG

P N D D S T N V A T I L * G S T Y Y G D

CCGAACGACGACAGTACTAACGTGGCCACGATCTTANNNGGGAGCACGTACTACGGTGAC

S V K E R F D I R R D * * * N T V T L S

TCCGTCAAAGAGCGCTTCGATATCCGTCGCGACNNNNNNNNNNAACACCGTTACCTTATCG

M D D L Q P E D S A E Y N C A G S * Y

ATGGACGATCTGCAACCGGAAGACTCTGCAGAATACAATTGTGCAGGTTCT(NNN)XTAC

H Y R G Q G T D V T V S S

CACTACCGTGGTCAGGGTACCGACGTTACCGTCTCGTCG

The underlined regions indicate the four affinity regions as depicted in figure 1. Although variations in length can occur in all affinity regions, the fourth underlined affinity region can tolerate extensive length variations (1-more than 40 amino acids). In figure 6 an example of this primary scaffold is depicted which has been adapted for CDR regions with known antigen binding properties (affinity region 1,2 and 4). These affinity regions originate from lama derived VHH antibody fragment clones 1HCV, which can bind human chorionic gonadotropin (HCG), those of 1MEL, which can bind lysozym, and the CDR regions of 1BZQ, which have been shown to bind bovine RNase A. The depicted DNA sequences in

figure 6b have been ordered at Genset's. The synthetically produced DNA sequence are tested for their binding properties. Therefore the DNA fragments are cloned into a phage display vector as described in example 1. Phage production and panning experiments are also performed as described above with the exception that the coating of immunotubes is done with respectively bovine RNase A, HCG or lysozym. Binding phages are eluted and individual clones are tested for their sequence. In order to improve the scaffold, a mutational program is started as described in example 1, preferably using the dITP method.

Example 3.

DESIGN OF A MINIMAL PRIMARY SCAFFOLD

[0041] A minimal scaffold is designed according to the requirements and features as described in example 2. However now only four and five beta-elements are used in the scaffold (see figure 3). In the case of 4 beta-elements amino acids side chains of beta-elements 2, 3, 7 and 8 that are forming the mantle of the new scaffold need to be adjusted for a watery environment. The same is true for the side chains of amino acid in beta-elements 2, 4, 7 and 8 in the case of a 5 beta-elements containing scaffold. The Fc-epsilon receptor type alpha (structural code 1J88 at VAST) is used as a template for comparative modelling to design a new small scaffold consisting of 4 beta-elements. For the five beta-elements containing scaffold immunoglobulin killer receptor 2dl2 (VAST code 2DLI) is used as a structural template for modelling.

Example 4.

STRUCTURAL ANALYSIS OF SCAFFOLDS

[0042] In order to verify the crystal structure of scaffolds and isolated binding proteins generated via the above mentioned examples, crystallographic methods are applied. However, it is known that not all proteins can be crystallized (about 50%) and thus not all proteins can be analyzed this way. Because the core of the proteins should consist of beta-elements and beta-sheets, circular dichroism (CD) spectroscopy is used to test the presence of beta-structures.

Example 5.

IMPROVING PROPERTIES OF A SCAFFOLD FOR APPLICATIONS

[0043] For certain applications the properties of a scaffold need to be optimized. For example heat stability, acid tolerance or proteolytic stability can be advantageous or even required in certain environments in order to function well. A mutational and re-selection program can be applied to create a new scaffold with similar binding properties but with improved properties. In this example a selected binding protein is improved to resist proteolytic degradation in a proteolytic environment. A mixture or a cascade of proteases or the environment of the application is used to select for stability. First the selected scaffold is mutated using the above mentioned dITP methods (example 1) or other mutational methods. Next a phage display library is build from the mutated PCR products so that the new scaffolds are expressed on the outside of phages. Next the phages are added to a the desired proteolytic active environment for a certain time at the future application temperature. The remaining phages are isolated and subjected to a panning procedure like described above (example 1). After extensive washing bound phages are eluted, infected in E.coli cells that bear F-pili and grown overnight on a agar plate containing the appropriate antibiotics. Individual clones are re-checked for their new properties and sequenced. The process of mutation introduction and selection can be repeated several times or other selection conditions can be applied in further optimization rounds.

Example 6: Random mutagenesis of multiple cloning sites (MCS) by PCR using dITP.

[0044] The MCS of vector pBluescript KS+ (Stratagene) is amplified with the T7 (5' AATACGACTCACTATAG 3') and the T3 (5' ATTAACCCTCACTAAAG 3') primer (Isogen?)in the following polymerase chain reaction (PCR): 50 ng of vector DNA is amplified by Taq DNA polymerase (0.3 units per reaction, Promega)) in a 50 ml reaction containing 50 ng of T3 and T7 primer, 2 mM MgCl2, 10mM Tris-HCl (pH 8.3), 50 mM KCl, 200 µM dNTPs (Roche Diagnostics) and with or without 200 µM dITP (Boerhinger). After 30 cycles (45 sec 94 °C, 45 sec 44 °C and 45 sec 72 °C) (Primus 96^{plus}, MWG-Biotech) a part of the PCR products are digested with the endonucleases XbaI, PstI, EcoRI, Xho I (Gibco BRL; 20 units per reaction) and run on a 2.5 % agarose gel, while the rest of the PCR products are cloned into the vector pGEMTeasy (Promega) and sequenced. The PCR products produced in the absence of dITP are digested completely, while the PCR products produced in the presence of dITP are only partly digested, showing mutagenisation

of the recognition site of the endonuclease..

Example 7: Controlled random mutagenesis of multiple cloning sites (MCS) with a different number of PCR cycli in the presence of dITP.

[0045] Amplification of the MCS of vector pBluescript KS+ is performed as in example 6 with the following modifications: the number of PCR cycles is varied from 5, 10, 20 to 30 cycles. The PCR products are treated as in example 6. The PCR products produced after 5 PCR cycles in the presence of dITP are nearly completely digested, but the PCR products produced after 30 cycles are clearly partly digested. The ratio between undigested and digested products increases with the number of PCR cycles. The different PCR products show an increase in the number of random mutations in the PCR products generated with a higher number of cycles.

Example 8: Controlled random mutagenesis of multiple cloning sites (MCS) with a different ratio of dNTPs and dITP.

[0046] Amplification of the MCS of vector pBluescript KS+ (stratagene) is performed as in example 6 with the following modifications: The concentration of dNTPs is varied from 20 μM, 40 μM, 100 μM to 200 μM and the dITP concentration is set at 200 μM. After 30 PCR cycles the products are digested as in example 6 and separated on a 2.5 % agarose gel. The higher the ratio dITP:dNTPs the more undigested product is present. The products show more mutations when generated with a higher ratio dITP:dNTPs.

Example 9: Controlled random mutagenesis of multiple cloning sites (MCS) with a different concentration of dITP.

[0047] Amplification of the MCS of vector pBluescript KS+ (stratagene) is performed as in example 6 with the following modifications: The concentration of dITP is varied from 0 μM, 100 μM to 200 μM and the dNTP concentration is set at either 100 μM. After 30 PCR cycles the products are digested as in example 6 and separated on a 2.5 % agarose gel. Undigested products are detected in the PCR products produced in the presence of dITP. The ratio between undigested versus digested products increases with a higher dITP concentration. The products show a higher number of mutations when produced in the presence of a higher concentration of dITP.

Example 10: Controlled random mutagenesis of multiple cloning sites (MCS) with dITP by using different DNA polymerases.

[0048] Amplification of the MCS of vector pBluescript KS+ (stratagene) is performed as in example 6 with the following modifications: in addition to the Taq DNA polymerase DNA polymerases with proofreading such as, pwo and Deep Vent are used for amplification. The PCR products are treated as in example 6. More undigested product is detected in the products generated with DNA polymerases having no proofreading activity. The PCR products show more mutations when generated by DNA polymerases with no proofreading activity.

Example 11: Controlled random mutagenesis of multiple cloning sites (MCS) with a different concentration of one of the dNTPs.

[0049] Amplification of the MCS of vector pBluescript KS+ (stratagene) is performed as in example 6 with the following modifications: four parallel PCR reactions are performed in which in each reaction the concentration of one of the dNTPs is lowered to 20 μM, while the concentration of the other three nucleotides is set 200 μM. The concentration of dITP is in all reactions 200 μM. The results show that preferably compatible nucleotides to the nucleotide present at a low concentration are mutated.

Example 12: Cloning of inosine containing PCR products

[0050] Products that contain substantial numbers of inosine residues appeared hard to clone. All tested DNA polymerases fail to extend if inosine residues are present in the template. Therefore inosine residues should be replaced or strands containing inosine residues should be copied otherwise. To overcome this problem we melted inosine containing templates in the presence of phosphorylated random oligo-mers (7-20 mers). Now at primers can anneal randomly over the inosine containing spot and thus a random nucleotide can be present opposite to the inosine residue. The addition of a non-proofreading polymerase, sufficient dNTP's, ATP and a DNA ligase can result in the formation of a template based DNA copy without inosines in it. After treatment with a proof reading DNA polymerase inosines are replaced with normal nucleotides. After amplification and cloning of the products high levels of mutations can be obtained. This whole process of inosine based mutations can be repeated with or without prior cloning steps until desired

levels of mutations are obtained.

Example 13: Random mutagenesis of CDR regions

[0051] Forward and reverse primers that anneal at -20 and +20 bp, respectively, around the CDR3 region of lama derived 1HCV, 1MEL and 1BZQ are used for amplification of this CDR region in the presence of dITP as described in example 6 and 12. These PCR products are cloned and sequenced as described in example 6. Random mutations are detected in the CDR3 region.

Example 14: Random mutagenesis of a mixture of CDR regions

[0052] Forward and reverse primers that anneal around CDR3 regions, such are used for amplification of these CDR regions in the presence of dITP as described in example 6 and 12. The PCR products are cloned and tested for binding properties as described in example 1,6 and 12. Random mutations are obviously present in at least part of all newly generated CDR3 regions.

Example 15: Random mutagenesis of scaffold

[0053] Primers annealing at the 5 prime and 3 prime end of the scaffold sequence are used for in the presence of dITP as described in example 6 and 12. These PCR products are cloned and sequenced as described in example 6 and 12. Random mutations are detected in the amplified and cloned fragments.

Description of the figures.

Figure 1:

Domain notification of immunoglobular structures.

[0054] The diagram represents a topology of a protein consisting of 9 beta-elements (indicated 1-9 from N-terminal to C-terminal). Beta elements 1,2, 6 and 7 (blocked line type) and elements 3,4,5,8 and 9 (straight line type) form two beta-sheets.

[0055] Eight loops (L1-L8) are responsible for the connection of all beta-elements. Loop 2, 4, 6 and 8 are located at the top site of the diagram and this represents the physical location of these loops in example proteins. The function of loops 2,4 and 8 in light and heavy chain variable domains is to bind antigens (marked with patterned regions). The position of L6 (also marked with a patterned region) also allows antigen binding activity, but it has not been reported yet as a region with such properties. Loops 1, 3,5 and 7 are located at the opposite site of the proteins.

Figure 2:

Schematic 3D topology of immunoglobular domains.

[0056] Nine beta-elements are depicted in the diagram. Each number corresponds to their position in the protein from N- to C-terminal (see figure 1). Both sets of elements 1, 2, 7 and 6 and 9, 8, 3, 4 and 5 form beta-sheets. The two beta-sheets are located in parallel to each other and are drawn from a top view perspective. The loops that connect the beta-elements are also depicted. Bold lines are connecting loops between beta-elements that are in top position while dashed lines indicate connecting loops that are located in bottom position.

Figure 3:

Schematic 3D-topology of scaffold domains.

[0057] Eight example topologies of protein structures that can be used for the presentation of antigen binding sites. The actual core of most scaffolds is formed by 4 beta-elements (2, 3, 7 and 8) arranged in two small beta-sheets formed by respectively elements 2 and 7 and by elements 8 and 3. The loops that connect the beta-elements are also depicted. Bold lines are connecting loops between beta-elements that are in top position while dashed lines indicate connecting loops that are located in bottom position. A connection that starts dashed and ends solid indicates a connection between a bottom and top part of beta-elements. The numbers of the beta-elements depicted in the diagram correspond to the numbers and positions mentioned in figures 1 and 2. Conserved loops are also indicated (e.g. L2 and L7) and also

correspond to the loop numbers in the other figures. Some 3D-structure topologies of putative scaffolds are also present in the natural repertoire. In such cases one example is given at the bottom of the mentioned figure.

Figure 4:

Modular Affinity & Scaffold Transfer (MAST) Technique.

[0058] Putative antigen binding proteins that contain a core structure as described here can be used for transfer operations. Also scaffold or core structures can be used for transfer actions. The transfer operation can occur between structural identical or comparable scaffolds or cores that differ in amino acid composition. Putative affinity regions can be transferred from one scaffold or core to another scaffold or core by for example PCR, restriction digestions, DNA synthesis or other molecular techniques. The results of such transfers is depicted here in a very schematic diagram. The putative (coding) binding regions from molecule A (top part, Affinity regions) and the scaffold (coding) region of molecule B (bottom part, framework regions) can be isolated by molecular means. After recombination of both elements (hybrid structure) a new molecule appears that has binding properties of molecule A and scaffold properties of scaffold B.

Figure 5:

Structural alignments of naturally occurring proteins that contain at least a part of an immunoglobular structure.

[0059] Using the VAST protocol (http://www.ncbi.nlm.nih.gov/Structure/VAST/vast.shtml) it is possible to find identical or partly identical structures. VAST uses an algorithm that can predicts the presence of identical elements in presubmitted structures with an example template structure. Here the structure of 1F2X, a functional camelid variable domain (VHH) containing an immunoglobular domain, was chosen as a template to identify natural occurring proteins in all kinds of organisms that share structural elements. Next the amino acid sequence of some of the retrieved proteins was aligned. The location of all 9 beta-elements organized in two beta-sheets is indicated by underlining. Some matched proteins contain all 9 beta-elements while others lack one or more of these. Despite the lack of one or more structural components, such proteins still form a basic common structure. Amino acid sequence comparisons show that there is hardly any conservation although the 3D-structures of these proteins is, at least partially, identical. Structural identical proteins from all kinds of organisms were retrieved this way, varying from bacteria to flies to human.

Figure 6:

Topology of a primary scaffold.

[0060]

a) Amino acid sequence of a primary scaffold that is used to obtain optimal (secondary) scaffolds. Amino acid sequences were obtained using modeling software (Modeller 6.0 and Insight II) in combination with DNA and proteins analysis software (Vector NTI suite 7.0, InforMax). Light and heavy chain variable regions were used as a template to design this primary scaffold. The numbers (1-9) indicate beta-elements and L1-L8 indicate loops similarly as described in earlier figures. Underlined regions (L2, L4, L6 and L8) indicate affinity regions located at one site of the proteins. L2, L4 and L8 correspond to the location of CDR regions.

b) Amino acid and corresponding DNA sequences of a primary scaffold in which example CDR regions are inserted. These sets of CDR regions, depicted underlined in the figure, originate from 3 different camelid derived VHH variable regions, known as 1BZQ, 1HCV and 1MEL.

References

[0061] Beste G, Schmidt FS, Stibora T, Skerra A. Proc Natl Acad Sci U S A. 1999 Mar Better M, Chang CP, Robinson RR, Horwitz AH. Science. 1988 May 20;240(4855):1041-3.

[0062] Davies J, Riechmann L. FEBS Lett. 1994 Feb 21;339(3):285-90.

[0063] Dimasi N, Martin F, Volpari C, Brunetti M, Biasiol G, Altamura S, Cortese R, De Francesco R, Steinkuhler C, Sollazzo M. J Virol. 1997 Oct;71(10):7461-9.

[0064] Hamers-Casterman C, Atarhouch T, Muyldermans S, Robinson G, Hamers C, Songa EB, Bendahman N, Hamers R. Nature. 1993 Jun 3;363(6428):446-8.

[0065] Lauwereys M, Arbabi Ghahroudi M, Desmyter A, Kinne J, Holzer W, De Genst E, Wyns L,Muyldermans S.

EMBO J. 1998 Jul 1;17(13):3512-20.

[0066] McConnell SJ, Hoess RH. J Mol Biol. 1995 Jul 21;250(4):460-70.

[0067] Smith GP, Patel SU, Windass JD, Thornton JM, Winter G, Griffiths AD. J Mol Biol. 1998 Mar 27;277(2):317-32.

[0068] Ku J, Schultz PG. Proc Natl Acad Sci U S A. 1995 Jul 3;92(14):6552-6. 2;96(5):1898-903.

[0069] Koide A, Bailey CW, Huang X, Koide S, J. Mol. Biol. 1998 284 (4): 1141-1151

[0070] Shusta EV, Pessi A, Bianchi E, Crameri A, Venturini S, Tramontano A, Sollazzo M. Nature. 1993 Mar 25;362 (6418):367-9.

[0071] Holler PD, Kieke MC, Kranz DM, Wittrup KD, Nat. Biotechnol. 2000 18(7): 754-759

[0072] Skerra A. Biochim Biophys Acta. 2000 Oct 18;1482(1-2):337-50. Review.

[0073] Skerra A. J Biotechnol. 2001 Jun;74(4):257-75.

[0074] Skerra A, Pluckthun A. Science 1988 May 20;240(4855):1038-41

SEQUENCE LISTING

<110> CatchMabs B.V.

<120> A structure for presenting desired peptide sequences

<130> P58644EP00

<140> 01204762.7
<141> 2001-12-10

<160> 22

<170> PatentIn Ver. 2.1

<210> 1
<211> 336
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Combined DNA/RNA Molecule: primary scaffold

<220>
<223> Description of Artificial Sequence: primary scaffold

<220>
<221> CDS
<222> (1)..(336)
<223> /note="Wherein N stands for unknown"

<400> 1

```
aat gtg aaa ctg gtt gaa aaa ggt ggc aat ttc gtc gaa aac gat gac   48
Asn Val Lys Leu Val Glu Lys Gly Gly Asn Phe Val Glu Asn Asp Asp
  1               5                  10                  15

gat ctt aag ctc acg tgc cgt gct gaa ggt nnn nnn nnn nnn nnn tac   96
Asp Leu Lys Leu Thr Cys Arg Ala Glu Gly Xaa Xaa Xaa Xaa Xaa Tyr
            20                  25                  30

tgc atg ggt tgg ttc cgt cag gcg ccg aac gac gac agt act aac gtg   144
Cys Met Gly Trp Phe Arg Gln Ala Pro Asn Asp Asp Ser Thr Asn Val
        35                  40                  45

gcc acg atc tta nnn ggg agc acg tac tac ggt gac tcc gtc aaa gag   192
Ala Thr Ile Leu Xaa Gly Ser Thr Tyr Tyr Gly Asp Ser Val Lys Glu
    50                  55                  60

cgc ttc gat atc cgt cgc gac nnn nnn nnn aac acc gtt acc tta tcg   240
Arg Phe Asp Ile Arg Arg Asp Xaa Xaa Xaa Asn Thr Val Thr Leu Ser
65                  70                  75                  80

atg gac gat ctg caa ccg gaa gac tct gca gaa tac aat tgt gca ggt   288
Met Asp Asp Leu Gln Pro Glu Asp Ser Ala Glu Tyr Asn Cys Ala Gly
                85                  90                  95

tct nnn tac cac tac cgt ggt cag ggt acc gac gtt acc gtc tcg tcg   336
Ser Xaa Tyr His Tyr Arg Gly Gln Gly Thr Asp Val Thr Val Ser Ser
            100                 105                 110
```

<210> 2
<211> 112
<212> PRT
<213> Artificial Sequence
<223> Description of Artificial Sequence: primary

scaffold

<400> 2
Asn Val Lys Leu Val Glu Lys Gly Gly Asn Phe Val Glu Asn Asp Asp
1 5 10 15

Asp Leu Lys Leu Thr Cys Arg Ala Glu Gly Xaa Xaa Xaa Xaa Xaa Tyr
20 25 30

Cys Met Gly Trp Phe Arg Gln Ala Pro Asn Asp Asp Ser Thr Asn Val
35 40 45

Ala Thr Ile Leu Xaa Gly Ser Thr Tyr Tyr Gly Asp Ser Val Lys Glu
50 55 60

Arg Phe Asp Ile Arg Arg Asp Xaa Xaa Xaa Asn Thr Val Thr Leu Ser
65 70 75 80

Met Asp Asp Leu Gln Pro Glu Asp Ser Ala Glu Tyr Asn Cys Ala Gly
85 90 95

Ser Xaa Tyr His Tyr Arg Gly Gln Gly Thr Asp Val Thr Val Ser Ser
100 105 110

<210> 3
<211> 17
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: primer T7

<220>
<221> misc_feature
<222> (1)..(17)

<400> 3
aatacgactc actatag 17

<210> 4
<211> 17
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: primer T3

<220>
<221> misc_feature
<222> (1)..(17)

<400> 4
attaaccctc actaaag 17

<210> 5
<211> 125
<212> PRT
<213> Lama

<220>
<221> SITE
<222> (1)..(125)
<223> /note="Antibody Cab-Ca05"

<400> 5
Gln Val Gln Leu Val Glu Ser Gly Gly Gly Ser Val Gln Ala Gly Gly
1 5 10 15

Ser Leu Arg Leu Ser Cys Ala Ala Ser Gly Tyr Thr Val Ser Thr Tyr
20 25 30

Cys Met Gly Trp Phe Arg Gln Ala Pro Gly Lys Glu Arg Glu Gly Val
35 40 45

Ala Thr Ile Leu Gly Gly Ser Thr Tyr Tyr Gly Asp Ser Val Lys Gly
50 55 60

Arg Phe Thr Ile Ser Gln Asp Asn Ala Lys Asn Thr Val Tyr Leu Gln
65 70 75 80

Met Asn Ser Leu Lys Pro Glu Asp Thr Ala Ile Tyr Tyr Cys Ala Gly
85 90 95

Ser Thr Val Ala Ser Thr Gly Trp Cys Ser Arg Leu Arg Pro Tyr Asp
100 105 110

Tyr His Tyr Arg Gly Gln Gly Thr Gln Val Thr Val Ser
115 120 125

<210> 6
<211> 127
<212> PRT
<213> Lama

<220>
<221> SITE
<222> (1)..(127)
<223> /note="Heavy chain variable domain"

<400> 6
Gln Val Gln Leu Gln Glu Ser Gly Gly Gly Leu Val Gln Ala Gly Gly
1 5 10 15

Ser Leu Arg Leu Ser Cys Ala Ala Ser Gly Arg Ala Ala Ser Gly His
20 25 30

Gly His Tyr Gly Met Gly Trp Phe Arg Gln Val Pro Gly Lys Glu Arg
35 40 45

Glu Phe Val Ala Ala Ile Arg Trp Ser Gly Lys Glu Thr Trp Tyr Lys
50 55 60

Asp Ser Val Lys Gly Arg Phe Thr Ile Ser Arg Asp Asn Ala Lys Thr
65 70 75 80

Thr Val Tyr Leu Gln Met Asn Ser Leu Lys Gly Glu Asp Thr Ala Val
85 90 95

Tyr Tyr Cys Ala Ala Arg Pro Val Arg Val Ala Asp Ile Ser Leu Pro
100 105 110

Val Gly Phe Asp Tyr Trp Gly Gln Gly Thr Gln Val Thr Val Ser
115 120 125

<210> 7
<211> 120
<212> PRT
<213> Homo sapiens

<220>

<221> SITE
<222> (1)..(120)
<223> /note="Igg1 heavy chain"

<400> 7
Ala Val Lys Leu Val Gln Ala Gly Gly Gly Val Val Gln Pro Gly Arg
1 5 10 15

Ser Leu Arg Leu Ser Cys Ile Ala Ser Gly Phe Thr Phe Ser Asn Tyr
20 25 30

Gly Met His Trp Val Arg Gln Ala Pro Gly Lys Gly Leu Glu Trp Val
35 40 45

Ala Val Ile Trp Tyr Asn Gly Ser Arg Thr Tyr Tyr Gly Asp Ser Val
50 55 60

Lys Gly Arg Phe Thr Ile Ser Arg Asp Asn Ser Lys Arg Thr Leu Tyr
65 70 75 80

Met Gln Met Asn Ser Leu Arg Thr Glu Asp Thr Ala Val Tyr Tyr Cys
85 90 95

Ala Arg Asp Pro Asp Ile Leu Thr Ala Phe Ser Phe Asp Tyr Trp Gly
100 105 110

Gln Gly Val Leu Val Thr Val Ser
115 120

<210> 8
<211> 89
<212> PRT
<213> Homo sapiens

<220>
<221> SITE
<222> (1)..(89)
<223> /note="Vcam-1"

<400> 8
Phe Lys Ile Glu Thr Thr Pro Glu Ser Arg Tyr Leu Ala Gln Ile Gly
1 5 10 15

Asp Ser Val Ser Leu Thr Cys Ser Thr Thr Gly Cys Glu Ser Pro Phe
20 25 30

Phe Ser Trp Arg Thr Gln Ile Asp Ser Pro Leu Asn Gly Lys Val Thr
35 40 45

Asn Glu Gly Thr Thr Ser Thr Leu Thr Met Asn Pro Val Ser Phe Gly
50 55 60

Asn Glu His Ser Tyr Leu Cys Thr Ala Thr Cys Glu Ser Arg Lys Leu
65 70 75 80

Glu Lys Gly Ile Gln Val Glu Ile Tyr
85

<210> 9
<211> 92
<212> PRT
<213> Hepatitis C virus

<220>
<221> SITE
<222> (1)..(92)

<223> /note="Hcv Protease"

<400> 9
Thr Gln Ser Phe Leu Ala Thr Cys Val Asn Gly Val Cys Trp Thr Val
1 5 10 15

Tyr His Gly Ala Gly Ser Lys Thr Leu Ala Gly Pro Lys Gly Pro Ile
20 25 30

Thr Gln Met Tyr Thr Asn Val Asp Gln Asp Leu Val Gly Trp Gln Ala
35 40 45

Pro Pro Gly Ala Arg Ser Leu Thr Pro Cys Thr Cys Gly Ser Ser Asp
50 55 60

Leu Tyr Leu Val Thr Arg His Ala Asp Val Ile Pro Val Arg Arg Arg
65 70 75 80

Gly Asp Ser Arg Gly Ser Leu Leu Ser Pro Arg Pro
85 90

<210> 10
<211> 102
<212> PRT
<213> Mus musculus

<220>
<221> SITE
<222> (1)..(102)
<223> /note="Soluble part of the Junction Adhesion
Molecule"

<400> 10
Lys Gly Ser Val Tyr Thr Ala Gln Ser Asp Val Gln Val Pro Glu Asn
1 5 10 15

Glu Ser Ile Lys Leu Thr Cys Thr Tyr Ser Gly Phe Ser Ser Pro Arg
20 25 30

Val Glu Trp Lys Phe Val Gln Gly Ser Thr Thr Ala Leu Val Cys Tyr
35 40 45

Asn Ser Gln Ile Thr Ala Pro Tyr Ala Asp Arg Val Thr Phe Ser Ser
50 55 60

Ser Gly Ile Thr Phe Ser Ser Val Thr Arg Lys Asp Asn Gly Glu Tyr
65 70 75 80

Thr Cys Met Val Ser Glu Glu Gly Gly Gln Asn Tyr Gly Glu Val Ser
85 90 95

Ile His Leu Thr Val Leu
100

<210> 11
<211> 91
<212> PRT
<213> Homo sapiens

<220>
<221> SITE
<222> (1)..(91)
<223> /note="Fragment of Fibronectin Encompassing
Type-Iii repeats 7 through 10"

<400> 11

```
Val Pro Pro Pro Thr Asp Leu Arg Phe Thr Asn Ile Gly Pro Asp Thr
 1               5                  10                  15

Met Arg Val Thr Trp Ala Pro Pro Pro Ser Ile Asp Leu Thr Asn Phe
            20                  25                  30

Leu Val Arg Tyr Ser Pro Val Lys Asn Glu Glu Asp Val Ala Glu Leu
        35                  40                  45

Ser Ile Ser Pro Ser Asp Asn Ala Val Val Leu Thr Asn Leu Leu Pro
    50                  55                  60

Gly Thr Glu Tyr Val Val Ser Val Ser Ser Val Tyr Glu Gln His Glu
65                  70                  75                  80

Ser Thr Pro Leu Arg Gly Arg Gln Lys Thr Gly
                85                  90
```

<210> 12
<211> 95
<212> PRT
<213> Drosophila

<220>
<221> SITE
<222> (1)..(95)
<223> /note="Neuroglian"

<400> 12

```
Pro Asn Ala Pro Lys Leu Thr Gly Ile Thr Cys Gln Ala Asp Lys Ala
 1               5                  10                  15

Glu Ile His Trp Glu Gln Gln Gly Asp Asn Arg Ser Pro Ile Leu His
            20                  25                  30

Tyr Thr Ile Gln Phe Asn Thr Ser Phe Thr Pro Ala Ser Trp Asp Ala
        35                  40                  45

Ala Tyr Glu Lys Val Pro Asn Thr Asp Ser Ser Phe Val Val Gln Met
    50                  55                  60

Ser Pro Trp Ala Asn Tyr Thr Phe Arg Val Ile Ala Phe Asn Lys Ile
65                  70                  75                  80

Gly Ala Ser Pro Pro Ser Ala His Ser Asp Ser Cys Thr Thr Gln
                85                  90                  95
```

<210> 13
<211> 100
<212> PRT
<213> Homo sapiens

<220>
<221> SITE
<222> (1)..(100)
<223> /note="Interleukin-4 receptor alpha chain complex"

<400> 13

```
Arg Ala Pro Gly Asn Leu Thr Val His Thr Asn Val Ser Asp Thr Leu
 1               5                  10                  15

Leu Leu Thr Trp Ser Asn Pro Tyr Pro Pro Asp Asn Tyr Leu Tyr Asn
            20                  25                  30

His Leu Thr Tyr Ala Val Asn Ile Ser Glu Asn Asp Pro Ala Asp Phe
        35                  40                  45
```

Arg Ile Tyr Asn Val Thr Tyr Leu Glu Pro Ser Leu Arg Ile Ala Ala
50 55 60

Ser Thr Leu Lys Ser Gly Ile Ser Tyr Arg Ala Arg Val Arg Ala Trp
65 70 75 80

Ala Gln Ala Tyr Asn Thr Thr Trp Ser Glu Trp Ser Pro Ser Thr Lys
85 90 95

Trp His Asn Ser
100

<210> 14
<211> 100
<212> PRT
<213> Escherichia coli

<220>
<221> SITE
<222> (1)..(100)
<223> /note="(Lacz) Beta-Galactosidase"

<400> 14
Phe Phe Gln Phe Arg Leu Ser Gly Gln Thr Ile Glu Val Thr Ser Glu
1 5 10 15

Tyr Leu Phe Arg His Ser Asp Asn Glu Leu Leu His Trp Met Val Ala
20 25 30

Leu Asp Gly Lys Pro Leu Ala Ser Gly Glu Val Pro Leu Asp Val Ala
35 40 45

Pro Gln Gly Lys Gln Leu Ile Glu Leu Pro Glu Leu Pro Gln Pro Glu
50 55 60

Ser Ala Gly Gln Leu Trp Leu Thr Val Arg Val Val Gln Pro Asn Ala
65 70 75 80

Thr Ala Trp Ser Glu Ala Gly His Ile Ser Ala Trp Gln Gln Trp Arg
85 90 95

Leu Ala Glu Asn
100

<210> 15
<211> 405
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Combined DNA/RNA Molecule: scaffold
with Vhh 1MEL CDR regions

<220>
<223> Description of Artificial Sequence: scaffold with
Vhh 1MEL CDR regions

<220>
<221> CDS
<222> (1)..(405)

<400> 15
aat gtg aaa ctg gtt gaa aaa ggt ggc aat ttc gtc gaa aac gat gac 48
Asn Val Lys Leu Val Glu Lys Gly Gly Asn Phe Val Glu Asn Asp Asp
1 5 10 15

```
gat ctt aag ctc acg tgc cgt gct gaa ggt tac acc att ggc ccg tac  96
Asp Leu Lys Leu Thr Cys Arg Ala Glu Gly Tyr Thr Ile Gly Pro Tyr
            20                  25                  30

tgc atg ggt tgg ttc cgt cag gcg ccg aac gac gac agt act aac gtg  144
Cys Met Gly Trp Phe Arg Gln Ala Pro Asn Asp Asp Ser Thr Asn Val
        35                  40                  45

gcc acg atc aac atg ggt ggc ggt att acg tac tac ggt gac tcc gtc  192
Ala Thr Ile Asn Met Gly Gly Gly Ile Thr Tyr Tyr Gly Asp Ser Val
    50                  55                  60

aaa gag cgc ttc gat atc cgt cgc gac aac gcg tcc aac acc gtt acc  240
Lys Glu Arg Phe Asp Ile Arg Arg Asp Asn Ala Ser Asn Thr Val Thr
65                  70                  75                  80

tta tcg atg gac gat ctg caa ccg gaa gac tct gca gaa tac aat tgt  288
Leu Ser Met Asp Asp Leu Gln Pro Glu Asp Ser Ala Glu Tyr Asn Cys
                85                  90                  95

gca ggt gat tct acc att tac gcg agc tat tat gaa tgt ggt cat ggc  336
Ala Gly Asp Ser Thr Ile Tyr Ala Ser Tyr Tyr Glu Cys Gly His Gly
            100                 105                 110

ctg agt acc ggc ggt tac ggc tac gat agc cac tac cgt ggt cag ggt  384
Leu Ser Thr Gly Gly Tyr Gly Tyr Asp Ser His Tyr Arg Gly Gln Gly
        115                 120                 125

acc gac gtt acc gtc tcg tcg                                     405
Thr Asp Val Thr Val Ser Ser
    130                 135
```

<210> 16
<211> 135
<212> PRT
<213> Artificial Sequence
<223> Description of Artificial Sequence: scaffold with Vhh 1MEL CDR regions

<400> 16

```
Asn Val Lys Leu Val Glu Lys Gly Gly Asn Phe Val Glu Asn Asp Asp
  1               5                  10                  15

Asp Leu Lys Leu Thr Cys Arg Ala Glu Gly Tyr Thr Ile Gly Pro Tyr
            20                  25                  30

Cys Met Gly Trp Phe Arg Gln Ala Pro Asn Asp Asp Ser Thr Asn Val
        35                  40                  45

Ala Thr Ile Asn Met Gly Gly Gly Ile Thr Tyr Tyr Gly Asp Ser Val
    50                  55                  60

Lys Glu Arg Phe Asp Ile Arg Arg Asp Asn Ala Ser Asn Thr Val Thr
65                  70                  75                  80

Leu Ser Met Asp Asp Leu Gln Pro Glu Asp Ser Ala Glu Tyr Asn Cys
                85                  90                  95

Ala Gly Asp Ser Thr Ile Tyr Ala Ser Tyr Tyr Glu Cys Gly His Gly
            100                 105                 110

Leu Ser Thr Gly Gly Tyr Gly Tyr Asp Ser His Tyr Arg Gly Gln Gly
        115                 120                 125

Thr Asp Val Thr Val Ser Ser
    130                 135
```

<210> 17
<211> 363
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Combined DNA/RNA Molecule: scaffold with Vhh 1BZQ CDR regions

<220>
<223> Description of Artificial Sequence: scaffold with Vhh 1BZQ CDR regions

<220>
<221> CDS
<222> (1)..(363)

<400> 17

```
aat gtg aaa ctg gtt gaa aaa ggt ggc aat ttc gtc gaa aac gat gac   48
Asn Val Lys Leu Val Glu Lys Gly Gly Asn Phe Val Glu Asn Asp Asp
  1               5                  10                  15

gat ctt aag ctc acg tgc cgt gct agc ggt tac gcc tac acg tat atc   96
Asp Leu Lys Leu Thr Cys Arg Ala Ser Gly Tyr Ala Tyr Thr Tyr Ile
            20                  25                  30

tac atg ggt tgg ttc cgt cag gcg ccg aac gac gac agt act aac gtg   144
Tyr Met Gly Trp Phe Arg Gln Ala Pro Asn Asp Asp Ser Thr Asn Val
        35                  40                  45

gcc acc atc gac tcg ggt ggc ggc ggt acc ctg tac ggt gac tcc gtc   192
Ala Thr Ile Asp Ser Gly Gly Gly Gly Thr Leu Tyr Gly Asp Ser Val
    50                  55                  60

aaa gag cgc ttc gat atc cgt cgc gac aaa ggc tcc aac acc gtt acc   240
Lys Glu Arg Phe Asp Ile Arg Arg Asp Lys Gly Ser Asn Thr Val Thr
65                  70                  75                  80

tta tcg atg gac gat ctg caa ccg gaa gac tct gca gaa tac aat tgt   288
Leu Ser Met Asp Asp Leu Gln Pro Glu Asp Ser Ala Glu Tyr Asn Cys
                85                  90                  95

gca gcg ggt ggc tac gaa ctg cgc gac cgc acc tac ggt cag cgt ggt   336
Ala Ala Gly Gly Tyr Glu Leu Arg Asp Arg Thr Tyr Gly Gln Arg Gly
            100                 105                 110

cag ggt acc gac gtt acc gtc tcg tcg                              363
Gln Gly Thr Asp Val Thr Val Ser Ser
        115                 120
```

<210> 18
<211> 121
<212> PRT
<213> Artificial Sequence
<223> Description of Artificial Sequence: scaffold with Vhh 1BZQ CDR regions

<400> 18

```
Asn Val Lys Leu Val Glu Lys Gly Gly Asn Phe Val Glu Asn Asp Asp
  1               5                  10                  15

Asp Leu Lys Leu Thr Cys Arg Ala Ser Gly Tyr Ala Tyr Thr Tyr Ile
            20                  25                  30
```

```
Tyr Met Gly Trp Phe Arg Gln Ala Pro Asn Asp Asp Ser Thr Asn Val
        35                  40                  45

Ala Thr Ile Asp Ser Gly Gly Gly Gly Thr Leu Tyr Gly Asp Ser Val
    50                  55                  60

Lys Glu Arg Phe Asp Ile Arg Arg Asp Lys Gly Ser Asn Thr Val Thr
65                  70                  75                  80

Leu Ser Met Asp Asp Leu Gln Pro Glu Asp Ser Ala Glu Tyr Asn Cys
                85                  90                  95

Ala Ala Gly Gly Tyr Glu Leu Arg Asp Arg Thr Tyr Gly Gln Arg Gly
            100                 105                 110

Gln Gly Thr Asp Val Thr Val Ser Ser
        115                 120
```

<210> 19
<211> 351
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Combined DNA/RNA Molecule: scaffold with Vhh 1HCV CDR regions

<220>
<223> Description of Artificial Sequence: scaffold with Vhh 1HCV CDR regions

<220>
<221> CDS
<222> (1)..(351)

<400> 19

```
aat gtg aaa ctg gtt gaa aaa ggt ggc aat ttc gtc gaa aac gat gac   48
Asn Val Lys Leu Val Glu Lys Gly Gly Asn Phe Val Glu Asn Asp Asp
  1               5                  10                  15

gat ctt aag ctc acg tgc cgt gct gaa ggt cgt acg ggt tcg acc tac   96
Asp Leu Lys Leu Thr Cys Arg Ala Glu Gly Arg Thr Gly Ser Thr Tyr
            20                  25                  30

gat atg ggt tgg ttc cgt cag gcg ccg aac gac gac agt act aac gtg   144
Asp Met Gly Trp Phe Arg Gln Ala Pro Asn Asp Asp Ser Thr Asn Val
        35                  40                  45

gcc acg atc aac tgg gat agc gcc cgt acg tac tac ggt gac tcc gtc   192
Ala Thr Ile Asn Trp Asp Ser Ala Arg Thr Tyr Tyr Gly Asp Ser Val
    50                  55                  60

aaa gag cgc ttc gat atc cgt cgc gac aat gcc tcc aac acc gtt acc   240
Lys Glu Arg Phe Asp Ile Arg Arg Asp Asn Ala Ser Asn Thr Val Thr
65                  70                  75                  80

tta tcg atg gac gat ctg caa ccg gaa gac tct gca gaa tac aat tgt   288
Leu Ser Met Asp Asp Leu Gln Pro Glu Asp Ser Ala Glu Tyr Asn Cys
                85                  90                  95

gca ggt ggt gaa ggc ggc acc tgg gat agc cgt ggt cag ggt acc gac   336
Ala Gly Gly Glu Gly Gly Thr Trp Asp Ser Arg Gly Gln Gly Thr Asp
            100                 105                 110

gtt acc gtc tcg tcg                                              351
Val Thr Val Ser Ser
```

115

<210> 20
<211> 117
<212> PRT
<213> Artificial Sequence
<223> Description of Artificial Sequence: scaffold with Vhh 1HCV CDR regions

<400> 20

```
Asn Val Lys Leu Val Glu Lys Gly Gly Asn Phe Val Glu Asn Asp Asp
 1               5                  10                  15

Asp Leu Lys Leu Thr Cys Arg Ala Glu Gly Arg Thr Gly Ser Thr Tyr
            20                  25                  30

Asp Met Gly Trp Phe Arg Gln Ala Pro Asn Asp Asp Ser Thr Asn Val
        35                  40                  45

Ala Thr Ile Asn Trp Asp Ser Ala Arg Thr Tyr Tyr Gly Asp Ser Val
    50                  55                  60

Lys Glu Arg Phe Asp Ile Arg Arg Asp Asn Ala Ser Asn Thr Val Thr
65                  70                  75                  80

Leu Ser Met Asp Asp Leu Gln Pro Glu Asp Ser Ala Glu Tyr Asn Cys
                85                  90                  95

Ala Gly Gly Glu Gly Gly Thr Trp Asp Ser Arg Gly Gln Gly Thr Asp
            100                 105                 110

Val Thr Val Ser Ser
        115
```

<210> 21
<211> 363
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Combined DNA/RNA Molecule: primary scaffold

<220>
<223> Description of Artificial Sequence: primary scaffold

<220>
<221> CDS
<222> (1)..(363)
<223> /note="N stands for unknown"

<400> 21

```
aat gtg aaa ctg gtt gaa aaa ggt ggc aat ttc gtc gaa aac gat gac   48
Asn Val Lys Leu Val Glu Lys Gly Gly Asn Phe Val Glu Asn Asp Asp
 1               5                  10                  15

gat ctt aag ctc acg tgc cgt gct nnn nnn nnn nnn nnn nnn nnn nnn   96
Asp Leu Lys Leu Thr Cys Arg Ala Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa
            20                  25                  30

nnn atg ggt tgg ttc cgt cag gcg ccg aac gac gac agt act aac gtg   144
Xaa Met Gly Trp Phe Arg Gln Ala Pro Asn Asp Asp Ser Thr Asn Val
        35                  40                  45
```

```
gcc acc atc gac nnn nnn nnn nnn nnn nnn tac ggt gac tcc gtc   192
Ala Thr Ile Asp Xaa Xaa Xaa Xaa Xaa Xaa Tyr Gly Asp Ser Val
     50                  55                  60

aaa gag cgc ttc gat atc cgt cgc gac aaa ggc tcc aac acc gtt acc   240
Lys Glu Arg Phe Asp Ile Arg Arg Asp Lys Gly Ser Asn Thr Val Thr
 65                  70                  75                  80

tta tcg atg gac gat ctg caa ccg gaa gac tct gca gaa tac aat tgt   288
Leu Ser Met Asp Asp Leu Gln Pro Glu Asp Ser Ala Glu Tyr Asn Cys
                 85                  90                  95

gca nnn nnn nnn nnn nnn nnn nnn nnn nnn nnn nnn nnn nnn nnn ggt   336
Ala Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Gly
            100                 105                 110

cag ggt acc gac gtt acc gtc tcg tcg                              363
Gln Gly Thr Asp Val Thr Val Ser Ser
        115                 120
```

<210> 22
<211> 121
<212> PRT
<213> Artificial Sequence
<223> Description of Artificial Sequence: primary scaffold

<400> 22

```
Asn Val Lys Leu Val Glu Lys Gly Gly Asn Phe Val Glu Asn Asp Asp
  1               5                  10                  15

Asp Leu Lys Leu Thr Cys Arg Ala Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa
             20                  25                  30

Xaa Met Gly Trp Phe Arg Gln Ala Pro Asn Asp Asp Ser Thr Asn Val
         35                  40                  45

Ala Thr Ile Asp Xaa Xaa Xaa Xaa Xaa Xaa Xaa Tyr Gly Asp Ser Val
     50                  55                  60

Lys Glu Arg Phe Asp Ile Arg Arg Asp Lys Gly Ser Asn Thr Val Thr
 65                  70                  75                  80

Leu Ser Met Asp Asp Leu Gln Pro Glu Asp Ser Ala Glu Tyr Asn Cys
                 85                  90                  95

Ala Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Gly
            100                 105                 110

Gln Gly Thr Asp Val Thr Val Ser Ser
        115                 120
```

Claims

1. A synthetic or recombinant proteinaceous molecule comprising a binding peptide and a core, said core comprising a b-barrel comprising at least 4 strands, wherein said b-barrel comprises at least two b-sheets, wherein each of said b-sheet comprises two of said strands and wherein said binding peptide is a peptide connecting two strands in said b-barrel and wherein said binding peptide is outside its natural context.

2. A proteinaceous molecule according to claim 1, wherein said b-barrel comprises at least 5 strands, wherein at least one of said sheets comprises 3 of said strands.

3. A proteinaceous molecule according to claim 1 or claim 2, wherein said b-barrel comprises at least 6 strands, wherein at least two of said sheets comprises 3 of said strands.

4. A proteinaceous molecule according to any one of claims 1-3, wherein said b-barrel comprises at least 7 strands, wherein at least one of said sheets comprises 4 of said strands.

5. A proteinaceous molecule according to any one of claims 1-4, wherein said b-barrel comprises at least 8 strands, wherein at least one of said sheets comprises 4 of said strands.

6. A proteinaceous molecule according to any one of claims 1-5, wherein said b-barrel comprises at least 9 strands, wherein at least one of said sheets comprises 4 of said strands.

7. A proteinaceous molecule according to any one of claims 1-6, wherein said binding peptide connects two strands of said b-barrel on the open side of said barrel.

8. A proteinaceous molecule according to any one of claims 1-7, wherein said binding peptide connects said at least two b-sheets of said barrel.

9. A proteinaceous molecule according to any one of claims 1-8, which comprises at least one further binding peptide.

10. A proteinaceous molecule according to any one of claims 1-9, which comprises three binding peptides and three connecting peptide sequences.

11. A proteinaceous molecule according to any one of claims 1-9, which comprises at least 4 binding peptides.

12. A proteinaceous molecule according to claim 11, wherein at least one binding peptide recognizes another target molecule than at least one of the other binding peptides.

13. A method for identifying a proteinaceous molecule with an altered binding property, comprising introducing an alteration in the core of proteinaceous molecules according to any one of claims 1-12, and selecting from said proteinaceous molecules, a proteinaceous molecule with an altered binding property.

14. A method for identifying a proteinaceous molecule with an altered structural property, comprising introducing an alteration in the core of proteinaceous molecules according to any one of claims 1-2, and selecting from said proteinaceous molecules, a proteinaceous molecule with an altered binding property.

15. A method according to claim 13 or 14, wherein said alteration comprises a post-translational modification.

16. A method according to any one of claims 13-15, wherein said alteration is introduced into a nucleic acid coding for said at least one proteinaceous molecule, the method further comprising expressing said nucleic acid in an expression system that is capable of producing said proteinaceous molecule.

17. A proteinaceous molecule obtainable by a method according to any one of claims 13-16.

18. A proteinaceous molecule according to any one of claims 1-12 or 17, which is derived from the immunoglobulin superfamily.

19. A proteinaceous molecule according to claim 18, wherein the exterior of the proteinaceous molecule is immunologically similar to the immunoglobulin superfamily molecule it was derived from.

20. A cell comprising a proteinaceous molecule according to any one of claims 1-12 or 17-19.

21. A method for producing a nucleic acid encoding a proteinacous molecule capable of displaying at least one desired peptide sequence comprising providing a nucleid acid sequence encoding at least a first and second structural region separated by a nucleic acid sequence encoding said desired peptide sequence or a region where such a sequence can be inserted and mutating said nucleic acid encoding said first and second structural regions to obtain a desired nucleic acid encoding said proteinacous molecule capable of displaying at least one desired peptide sequence.

22. A method for displaying a desired peptide sequence, providing a nucleic acid encoding at least a two b-sheets, said , said b-sheets forming a b-barrel, said nucleic acid comprising a region for inserting a sequence encoding said desired peptide sequence, inserting a nucleic acid sequence comprising a desired peptide sequence, and expressing said nucleid acid whereby said b sheets are obtainable by a method according to claim 21.

23. Use of a proteinaceous molecule according to any one of claims 1-12 or 17-19, for separating a substance from a mixture.

24. A use according to claim 23, wherein said mixture is a biological fluid.

25. A use according to claim 24, wherein said biological fluis is an excretion product of an organism.

26. A use according to claim 25, wherein said excretion product is milk or a derivative of milk.

27. A use according to claim 24, wherein said mixture is blood or a derivative thereof.

28. A proteinaceous molecule according to any one of claims 1-12 or 17-19, for use as a pharmaceutical.

29. Use of a proteinaceous molecule according to any one of claims 1-12 or 17-19, in the preparation of a pharmaceutical formulation for the treatment of a pathological condition involving unwanted proteins or cells or microorganisms.

30. Use of a proteinaceous molecule according to any one of claims 1-12 or 17-19, in the preparation of a diagnostic assay.

31. A gene delivery vehicle comprising a proteinaceous molecule according to any one of claims 1-12 or 17-19. and a gene of interest.

32. A gene delivery vehicle comprising a nucleic acid encoding proteinaceous molecule according to any one of claims 1-12 or 17-19. and a nucleic acid sequence encoding a gene of interest.

33. A proteinaceous molecule according to any one of claims 1-12 or 17-19 conjugated to a moiety of interest.

34. A proteinaceous molecule according to claim 33, wherein said moiety of interest is a toxic moiety.

35. A chromatography column comprising a proteinaceous molecule according to any one of claims 1-12 or 17-19 and a packing material.

36. A nucleic acid obtainable by the method of claim 21.

37. A nucleic acid library comprising a collection of different nucleic acids according to claim 36.

38. A nucleic acid library according to claim 37, further comprising a collection of nucleic acids encoding different affinity regions.

39. A library according to claim 37 or 38, which is an expression library.

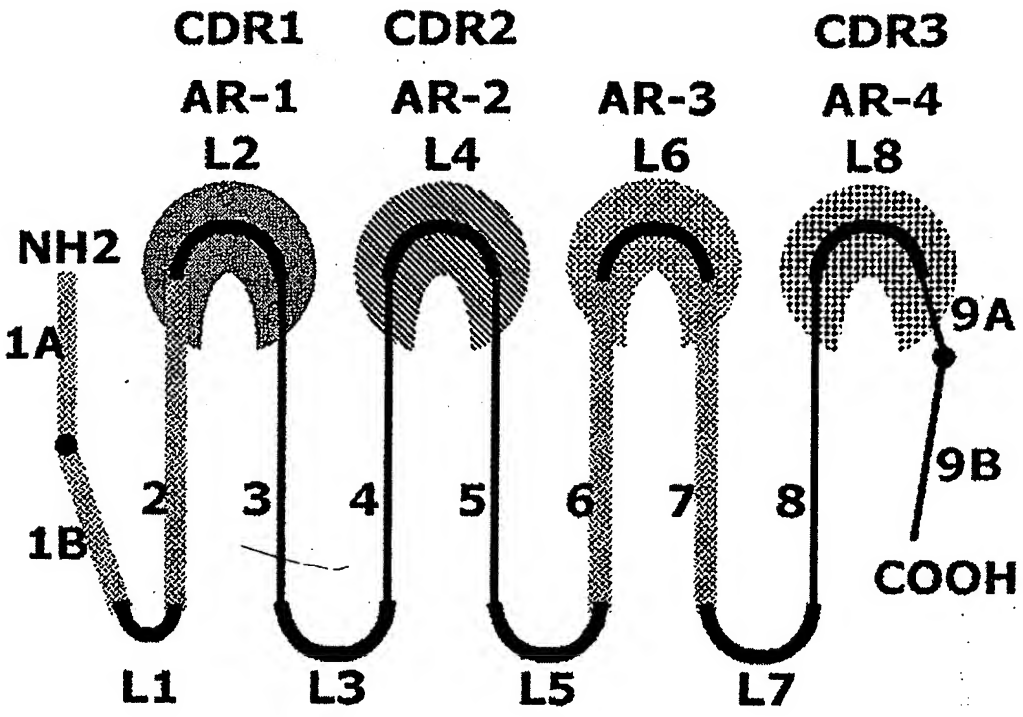

CDR1
CDR2
CDR3
AR-1
AR-2
AR-3
AR-4
L2
L4
L6
L8
NH2
1A
9A
9B
1B
2
3
4
5
6
7
8
COOH
L1
L3
L5
L7

Fig. 1

3D structure topology

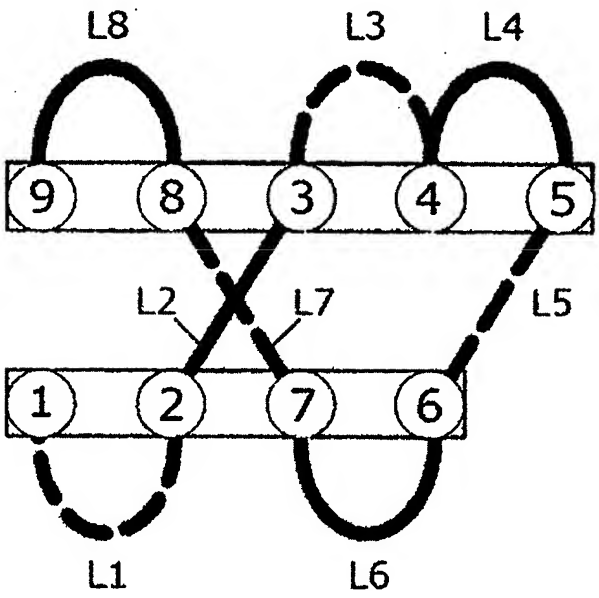
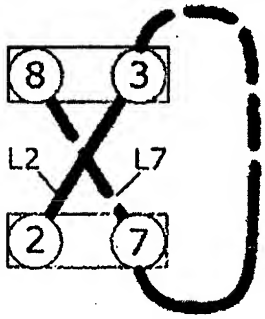


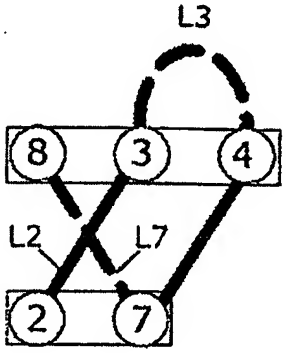
Fig. 2

Structural Deviations

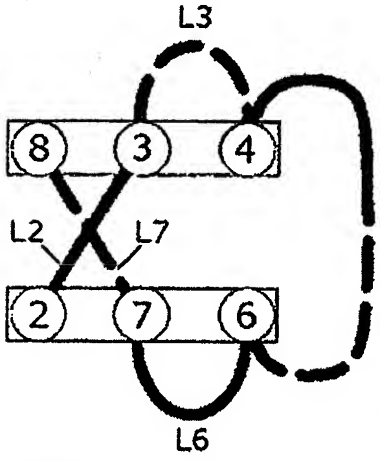
4 beta elements:



5 beta elements:

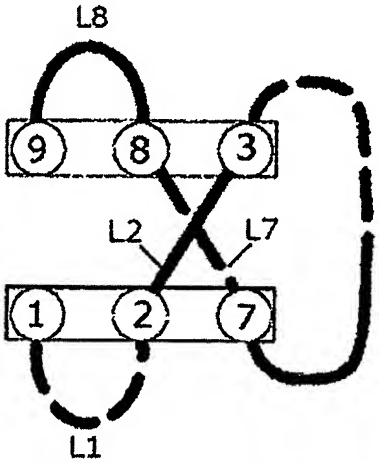


6 beta elements-a:



1GOY: Interleukin-1 receptor type 1

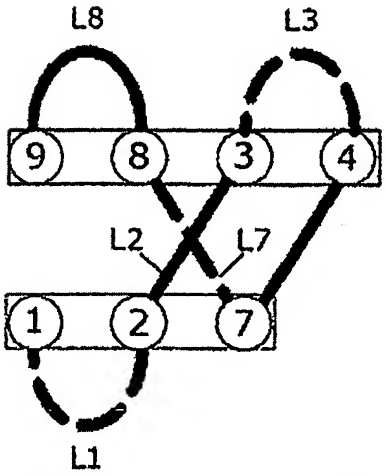
6 beta elements-b:



1J88: Fc epsilon receptor type alpha

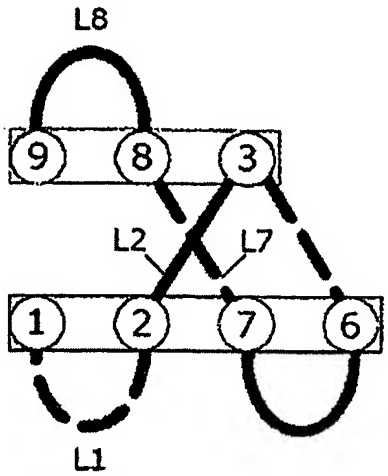
Fig. 3a

7 beta elements-a:



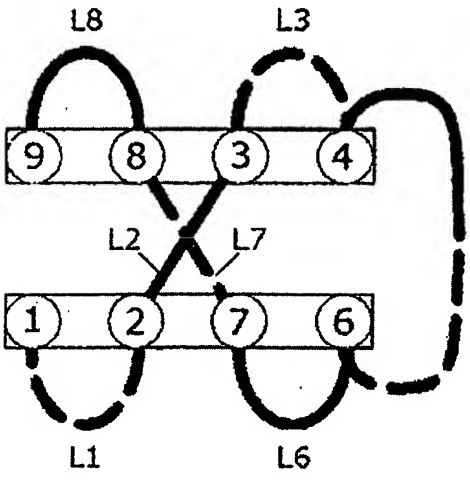
2DLI: Immunoglobulin killer receptor 2dl2

7 beta elements-b:



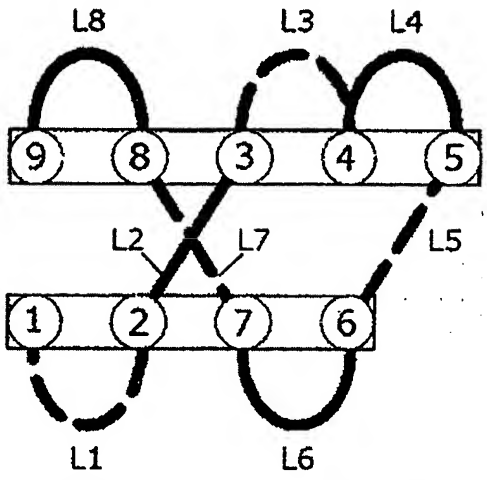
1FF5: E-cadherin domain

8 beta elements:



1IAR: Interleukin-4 alpha receptor

9 beta elements:



All antibody and T-cell receptor variable domains

Fig. 3b

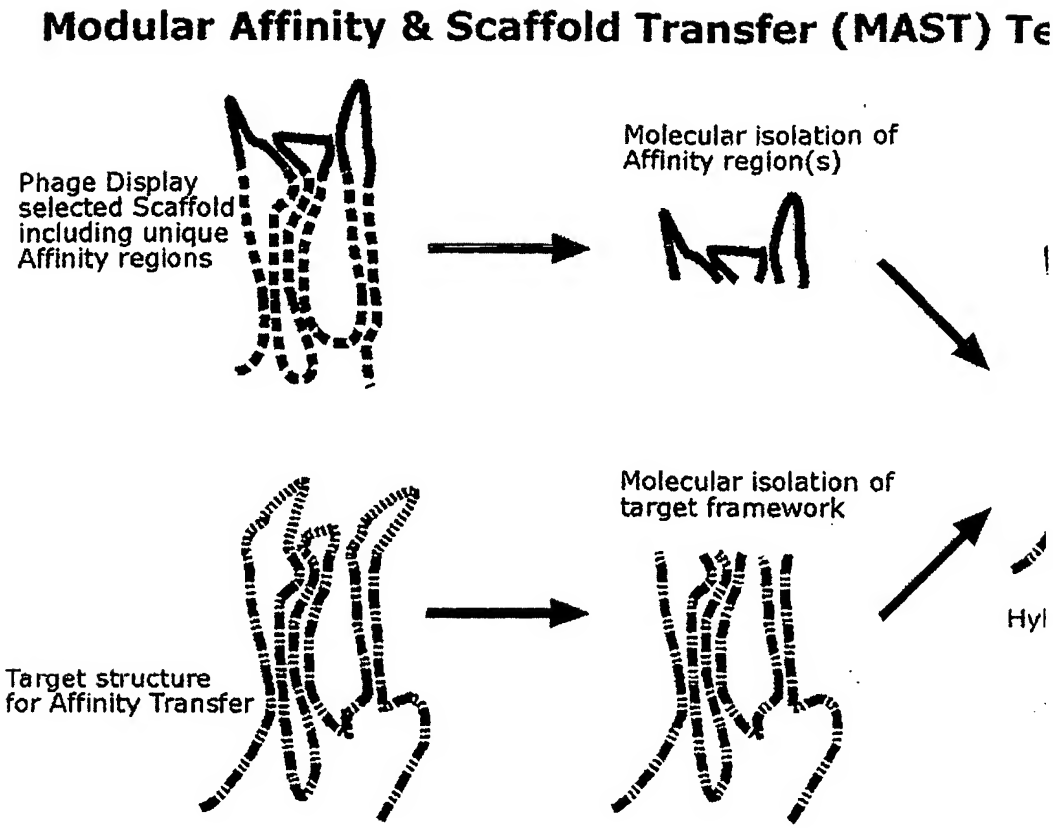

Modular Affinity & Scaffold Transfer (MAST) Te
Phage Display
selected Scaffold
including unique
Affinity regions
Molecular isolation of
Affinity region(s)
Molecular isolation of
target framework
Target structure
for Affinity Transfer
Hyl

Fig. 4

STRUCTURAL ALIGNMENT: EXAMPLES OF 1F2X V_{HH} CAMELID ANTIBODY VARIABLE FRAGMENT

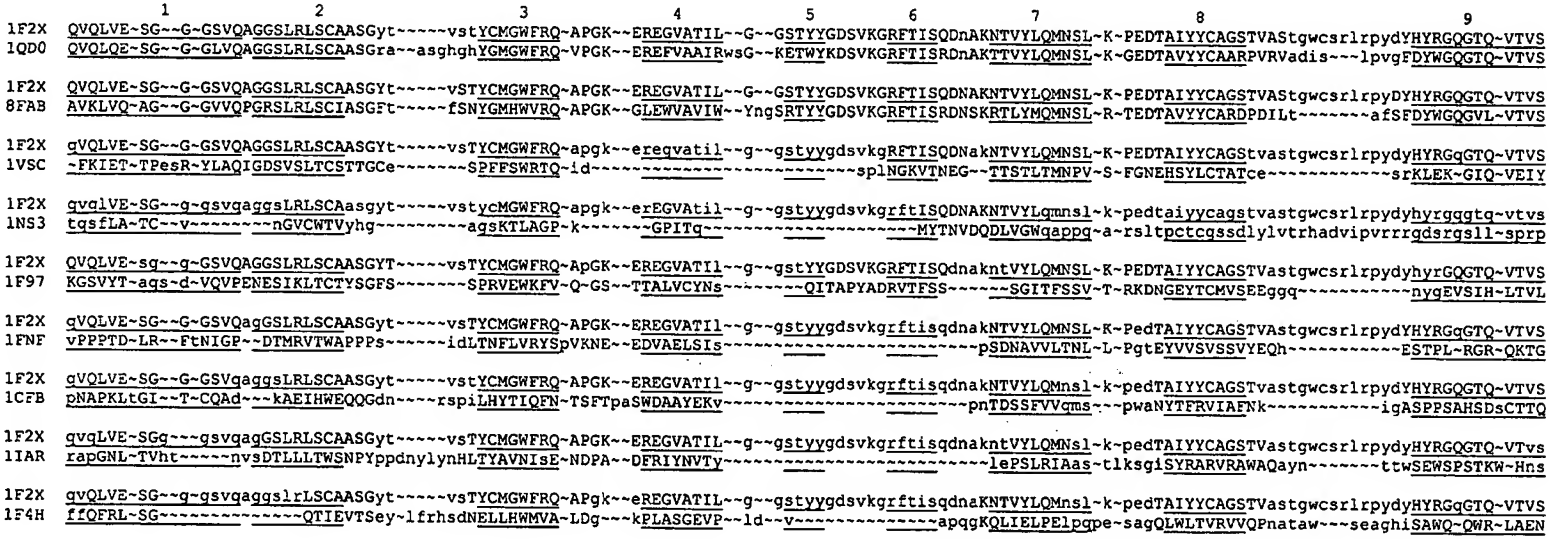


Figure elements explaned:
Underlined domains represent (putative) beta-elements 1-9
Capital letters indicate conserved structural amino acids
~ indicate the absence of amino acid residues

Aligned protein domains from:
1F2X Single domain camelid antibody Cab-Ca05
1QD0 Camelid heavy chain variable domain
8FAB Heavy chain from human Igg1
1VSC Human Vcam-1
1NS3 Structure Of Hcv Protease (Bk Strain)from hepatitis C-virus
1F97 Soluble Part Of The Junction Adhesion Molecule From Mouse
1FNF Fragment Of Human Fibronectin Encompassing Type-Iii Repeats 7 Through 10
1CFB Drosophila neuroglian
1IAR Human Interleukin-4 receptor alpha chain complex
1F4H E. Coli (Lacz) Beta-Galactosidase (Orthorhombic)

Fig. 5

EP 1 318 195 A1

Scaffold with V_{HH} 1MEL CDR regi ns

```
 N  V  K  L  V  E  K  G  G  N  F  V  E  N  D  D  D  L  K  L
AATGTGAAACTGGTTGAAAAAGGTGGCAATTTCGTCGAAAACGATGACGATCTTAAGCTC
 T  C  R  A  E  G  Y  T  I  G  P  Y  C  M  G  W  F  R  Q  A
ACGTGCCGTGCTGAAGGTTACACCATTGGCCCGTACTGCATGGGTTGGTTCCGTCAGGCG
 P  N  D  D  S  T  N  V  A  T  I  N  M  G  G  G  I  T  Y  Y
CCGAACGACGACAGTACTAACGTGGCCACGATCAACATGGGTGGCGGTATTACGTACTAC
 G  D  S  V  K  E  R  F  D  I  R  R  D  N  A  S  N  T  V  T
GGTGACTCCGTCAAAGAGCGCTTCGATATCCGTCGCGACAACGCGTCCAACACCGTTACC
 L  S  M  D  D  L  Q  P  E  D  S  A  E  Y  N  C  A  G  D  S
TTATCGATGGACGATCTGCAACCGGAAGACTCTGCAGAATACAATTGTGCAGGTGATTCT
 T  I  Y  A  S  Y  Y  E  C  G  H  G  L  S  T  G  G  Y  G  Y
ACCATTTACGCGAGCTATTATGAATGTGGTCATGGCCTGAGTACCGGCGGTTACGGCTAC
 D  S  H  Y  R  G  Q  G  T  D  V  T  V  S  S
GATAGCCACTACCGTGGTCAGGGTACCGACGTTACCGTCTCGTCG
```

Scaffold with V_{HH} 1BZQ CDR regions

```
 N  V  K  L  V  E  K  G  G  N  F  V  E  N  D  D  D  L  K  L
AATGTGAAACTGGTTGAAAAAGGTGGCAATTTCGTCGAAAACGATGACGATCTTAAGCTC
 T  C  R  A  S  G  Y  A  Y  T  Y  I  Y  M  G  W  F  R  Q  A
ACGTGCCGTGCTAGCGGTTACGCCTACACGTATATCTACATGGGTTGGTTCCGTCAGGCG
 P  N  D  D  S  T  N  V  A  T  I  D  S  G  G  G  G  T  L  Y
CCGAACGACGACAGTACTAACGTGGCCACCATCGACTCGGGTGGCGGCGGTACCCTGTAC
 G  D  S  V  K  E  R  F  D  I  R  R  D  K  G  S  N  T  V  T
GGTGACTCCGTCAAAGAGCGCTTCGATATCCGTCGCGACAAAGGCTCCAACACCGTTACC
 L  S  M  D  D  L  Q  P  E  D  S  A  E  Y  N  C  A  A  G  G
TTATCGATGGACGATCTGCAACCGGAAGACTCTGCAGAATACAATTGTGCAGCGGGTGGC
 Y  E  L  R  D  R  T  Y  G  Q  R  G  Q  G  T  D  V  T  V  S
TACGAACTGCGCGACCGCACCTACGGTCAGCGTGGTCAGGGTACCGACGTTACCGTCTCG
 S
TCG
```

Scaffold with V_{HH} 1HCV CDR regions

```
 N  V  K  L  V  E  K  G  G  N  F  V  E  N  D  D  D  L  K  L
AATGTGAAACTGGTTGAAAAAGGTGGCAATTTCGTCGAAAACGATGACGATCTTAAGCTC
 T  C  R  A  E  G  R  T  G  S  T  Y  D  M  G  W  F  R  Q  A
ACGTGCCGTGCTGAAGGTCGTACGGGTTCGACCTACGATATGGGTTGGTTCCGTCAGGCG
 P  N  D  D  S  T  N  V  A  T  I  N  W  D  S  A  R  T  Y  Y
CCGAACGACGACAGTACTAACGTGGCCACGATCAACTGGGATAGCGCCCGTACGTACTAC
 G  D  S  V  K  E  R  F  D  I  R  R  D  N  A  S  N  T  V  T
GGTGACTCCGTCAAAGAGCGCTTCGATATCCGTCGCGACAATGCCTCCAACACCGTTACC
 L  S  M  D  D  L  Q  P  E  D  S  A  E  Y  N  C  A  G  G  E
TTATCGATGGACGATCTGCAACCGGAAGACTCTGCAGAATACAATTGTGCAGGTGGTGAA
 G  G  T  W  D  S  R  G  Q  G  T  D  V  T  V  S  S
GGCGGCACCTGGGATAGCCGTGGTCAGGGTACCGACGTTACCGTCTCGTCG
```

Underlined regions indicate specific affinity regions.
The sequence of underlined regions in each panel represent respectively loop L2 (~CDR1 and AR1), L4 (~CDR2 and AR2) and L8 (~CDR3 and AR4).

Fig. 6a

Strucutural topology of a primairy scaffold

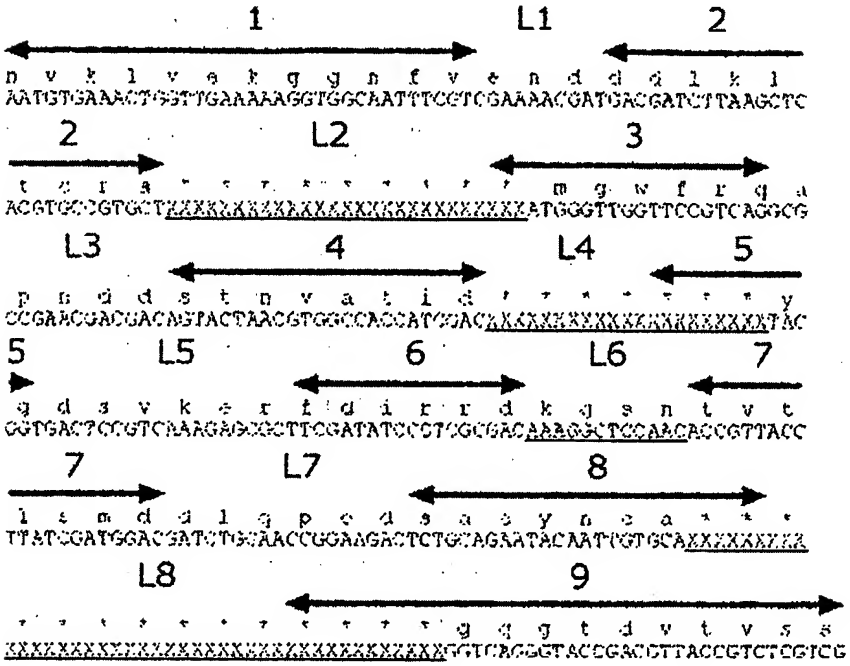


Fig. 6b

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
7 September 2001 (07.09.2001)

PCT

(10) International Publication Number
WO 01/64942 A1

(51) International Patent Classification[7]: C12Q 1/00, 1/70, 1/68, G01N 33/53

(21) International Application Number: PCT/US01/06414

(22) International Filing Date: 28 February 2001 (28.02.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
09/515,260 29 February 2000 (29.02.2000) US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier application:
US 09/515,260 (CIP)
Filed on 29 February 2000 (29.02.2000)

(71) Applicant *(for all designated States except US)*: **PHYLOS, INC.** [US/US]; 128 Spring Street, Lexington, MA 02421 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **LIPOVSEK, Dasa** [SI/US]; 45 Sunset Road, Cambridge, MA 02138 (US). **WAGNER, Richard, W.** [US/US]; 1007 Lowell Road, Concord, MA 01742 (US). **KUIMELIS, Robert, G.** [US/US]; 21 Malbert Road, Brighton, MA 02135 (US).

(74) Agent: **ELBING, Karen, L.**; Clark & Elbing LLP, 176 Federal Street, Boston, MA 02110-2214 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*
— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 01/64942 A1

(54) Title: PROTEIN SCAFFOLDS FOR ANTIBODY MIMICS AND OTHER BINDING PROTEINS

(57) Abstract: Disclosed herein are proteins that include a fibronectin type III domain having at least one randomized loop. Also disclosed herein are nucleic acids encoding such proteins and the use of such proteins in diagnostic methods and in methods for evolving novel compound-binding species and their ligands.

PROTEIN SCAFFOLDS FOR ANTIBODY MIMICS AND OTHER BINDING PROTEINS

Background of the Invention

This invention relates to protein scaffolds useful, for example, for the generation of products having novel binding characteristics.

Proteins having relatively defined three-dimensional structures, commonly referred to as protein scaffolds, may be used as reagents for the design of engineered products. These scaffolds typically contain one or more regions which are amenable to specific or random sequence variation, and such sequence randomization is often carried out to produce libraries of proteins from which desired products may be selected. One particular area in which such scaffolds are useful is the field of antibody design.

A number of previous approaches to the manipulation of the mammalian immune system to obtain reagents or drugs have been attempted. These have included injecting animals with antigens of interest to obtain mixtures of polyclonal antibodies reactive against specific antigens, production of monoclonal antibodies in hybridoma cell culture (Koehler and Milstein, Nature 256:495, 1975), modification of existing monoclonal antibodies to obtain new or optimized recognition properties, creation of novel antibody fragments with desirable binding characteristics, and randomization of single chain antibodies (created by connecting the variable regions of the heavy and light chains of antibody molecules with a flexible peptide linker) followed by selection for antigen binding by phage display (Clackson et al., Nature 352:624, 1991).

In addition, several non-immunoglobulin protein scaffolds have been proposed for obtaining proteins with novel binding properties. For example, a "minibody" scaffold, which is related to the immunoglobulin fold, has been designed by deleting three beta strands from a heavy chain variable domain of a monoclonal antibody (Tramontano et al., J. Mol. Recognit. 7:9, 1994). This protein includes 61 residues and can be used to present two hypervariable loops. These two loops have been randomized and products selected for antigen binding, but thus far the framework appears to have somewhat limited utility due to solubility problems. Another framework used to display loops has been tendamistat, a 74 residue, six-strand beta sheet sandwich held together by two disulfide bonds (McConnell and Hoess, J. Mol. Biol. 250:460, 1995). This scaffold includes three loops, but, to date, only two of these loops have been examined for randomization potential.

Other proteins have been tested as frameworks and have been used to display randomized residues on alpha helical surfaces (Nord et al., Nat. Biotechnol. 15:772, 1997; Nord et al., Protein Eng. 8:601, 1995), loops between alpha helices in alpha helix bundles (Ku and Schultz, Proc. Natl. Acad. Sci. USA 92:6552, 1995), and loops constrained by disulfide bridges, such as those of the small protease inhibitors (Markland et al., Biochemistry 35:8045, 1996; Markland et al., Biochemistry 35:8058, 1996; Rottgen and Collins, Gene 164:243, 1995; Wang et al., J. Biol. Chem. 270:12250, 1995).

Summary of the Invention

The present invention provides a new family of proteins capable of evolving to bind any compound of interest. These proteins, which make use of a fibronectin or fibronectin-like scaffold, function in a manner characteristic of natural or engineered antibodies (that is, polyclonal, monoclonal, or

single-chain antibodies) and, in addition, possess structural advantages. Specifically, the structure of these antibody mimics has been designed for optimal folding, stability, and solubility, even under conditions which normally lead to the loss of structure and function in antibodies.

These antibody mimics may be utilized for the purpose of designing proteins which are capable of binding to virtually any compound (for example, any protein) of interest. In particular, the fibronectin-based molecules described herein may be used as scaffolds which are subjected to directed evolution designed to randomize one or more of the three fibronectin loops which are analogous to the complementarity-determining regions (CDRs) of an antibody variable region. Such a directed evolution approach results in the production of antibody-like molecules with high affinities for antigens of interest. In addition, the scaffolds described herein may be used to display defined exposed loops (for example, loops previously randomized and selected on the basis of antigen binding) in order to direct the evolution of molecules that bind to such introduced loops. A selection of this type may be carried out to identify recognition molecules for any individual CDR-like loop or, alternatively, for the recognition of two or all three CDR-like loops combined into a non-linear epitope.

Accordingly, the present invention features a protein that includes a fibronectin type III domain having at least one randomized loop, the protein being characterized by its ability to bind to a compound that is not bound by the corresponding naturally-occurring fibronectin.

In preferred embodiments, the fibronectin type III domain is a mammalian (for example, a human) fibronectin type III domain; and the protein includes the tenth module of the fibronectin type III (^{10}Fn3) domain. In such proteins, compound binding is preferably mediated by either one, two, or

three ^{10}Fn3 loops. In other preferred embodiments, the second loop of ^{10}Fn3 may be extended in length relative to the naturally-occurring module, or the ^{10}Fn3 may lack an integrin-binding motif. In these molecules, the integrin-binding motif may be replaced by an amino acid sequence in which a basic amino acid-neutral amino acid-acidic amino acid sequence (in the N-terminal to C-terminal direction) replaces the integrin-binding motif; one preferred sequence is serine-glycine-glutamate. In another preferred embodiment, the fibronectin type III domain-containing proteins of the invention lack disulfide bonds.

Any of the fibronectin type III domain-containing proteins described herein may be formulated as part of a fusion protein (for example, a fusion protein which further includes an immunoglobulin F_c domain, a complement protein, a toxin protein, or an albumin protein). In addition, any of the fibronectin type III domain proteins may be covalently bound to a nucleic acid (for example, an RNA), and the nucleic acid may encode the protein. Moreover, the protein may be a multimer, or, particularly if it lacks an integrin-binding motif, it may be formulated in a physiologically-acceptable carrier.

The present invention also features proteins that include a fibronectin type III domain having at least one mutation in a β-sheet sequence which changes the scaffold structure. Again, these proteins are characterized by their ability to bind to compounds that are not bound by the corresponding naturally-occurring fibronectin.

In addition, any of the fibronectin scaffolds of the invention may be immobilized on a solid support (for example, a bead or chip), and these scaffolds may be arranged in any configuration on the solid support, including an array.

In a related aspect, the invention further features nucleic acids encoding any of the proteins of the invention. In preferred embodiments, the nucleic acid is DNA or RNA.

In another related aspect, the invention also features a method for generating a protein which includes a fibronectin type III domain and which is pharmaceutically acceptable to a mammal, involving removing the integrin-binding domain of said fibronectin type III domain. This method may be applied to any of the fibronectin type III domain-containing proteins described above and is particularly useful for generating proteins for human therapeutic applications. The invention also features such fibronectin type III domain-containing proteins which lack integrin-binding domains.

In yet other related aspects, the invention features screening methods which may be used to obtain or evolve randomized fibronectin type III proteins capable of binding to compounds of interest, or to obtain or evolve compounds (for example, proteins) capable of binding to a particular protein containing a randomized fibronectin type III motif. In addition, the invention features screening procedures which combine these two methods, in any order, to obtain either compounds or proteins of interest.

In particular, the first screening method, useful for the isolation or identification of randomized proteins of interest, involves: (a) contacting the compound with a candidate protein, the candidate protein including a fibronectin type III domain having at least one randomized loop, the contacting being carried out under conditions that allow compound-protein complex formation; and (b) obtaining, from the complex, the protein which binds to the compound.

The second screening method, for isolating or identifying a compound which binds to a protein having a randomized fibronectin type III domain, involves: (a) contacting the protein with a candidate compound, the contacting being carried out under conditions that allow compound-protein complex formation; and (b) obtaining, from the complex, the compound which binds to the protein.

In preferred embodiments, the methods further involve either randomizing at least one loop of the fibronectin type III domain of the protein obtained in step (b) and repeating steps (a) and (b) using the further randomized protein, or modifying the compound obtained in step (b) and repeating steps (a) and (b) using the further modified compound. In addition, the compound is preferably a protein, and the fibronectin type III domain is preferably a mammalian (for example, a human) fibronectin type III domain. In other preferred embodiments, the protein includes the tenth module of the fibronectin type III domain (^{10}Fn3), and binding is mediated by one, two, or three ^{10}Fn3 loops. In addition, the second loop of ^{10}Fn3 may be extended in length relative to the naturally-occurring module, or ^{10}Fn3 may lack an integrin-binding motif. Again, as described above, the integrin-binding motif may be replaced by an amino acid sequence in which a basic amino acid-neutral amino acid-acidic amino acid sequence (in the N-terminal to C-terminal direction) replaces the integrin-binding motif; one preferred sequence is serine-glycine-glutamate.

The selection methods described herein may be carried out using any fibronectin type III domain-containing protein. For example, the fibronectin type III domain-containing protein may lack disulfide bonds, or may be formulated as part of a fusion protein (for example, a fusion protein which further includes an immunoglobulin F_c domain, a complement protein, a toxin

protein, or an albumin protein). In addition, selections may be carried out using the fibronectin type III domain proteins covalently bound to nucleic acids (for example, RNAs or any nucleic acid which encodes the protein). Moreover, the selections may be carried out using fibronectin domain-containing protein multimers.

Preferably, the selections involve the immobilization of the binding target on a solid support. Preferred solid supports include columns (for example, affinity columns, such as agarose columns) or microchips.

In addition, the invention features diagnostic methods which employ the fibronectin scaffold proteins of the invention. Such diagnostic methods may be carried out on a sample (for example, a biological sample) to detect one analyte or to simultaneously detect many different analytes in the sample. The method may employ any of the scaffold molecules described herein. Preferably, the method involves (a) contacting the sample with a protein which binds to the compound analyte and which includes a fibronectin type III domain having at least one randomized loop, the contacting being carried out under conditions that allow compound-protein complex formation; and (b) detecting the complex, and therefore the compound in the sample.

In preferred embodiments, the protein is immobilized on a solid support (for example, a chip or bead) and may be immobilized as part of an array. The protein may be covalently bound to a nucleic acid, preferably, a nucleic acid, such as RNA, that encodes the protein. In addition, the compound is often a protein, but may also be any other analyte in a sample. Detection may be accomplished by any standard technique including, without limitation, radiography, fluorescence detection, mass spectroscopy, or surface plasmon resonance.

As used herein, by "fibronectin type III domain" is meant a domain having 7 or 8 beta strands which are distributed between two beta sheets, which themselves pack against each other to form the core of the protein, and further containing loops which connect the beta strands to each other and are solvent exposed. There are at least three such loops at each edge of the beta sheet sandwich, where the edge is the boundary of the protein perpendicular to the direction of the beta strands. Preferably, a fibronectin type III domain includes a sequence which exhibits at least 30% amino acid identity, and preferably at least 50% amino acid identity, to the sequence encoding the structure of the ^{10}Fn3 domain referred to as "1ttg" (ID = "1ttg" (one ttg)) available from the Protein Data Base. Sequence identity referred to in this definition is determined by the Homology program, available from Molecular Simulation (San Diego, CA). The invention further includes polymers of ^{10}Fn3-related molecules, which are an extension of the use of the monomer structure, whether or not the subunits of the polyprotein are identical or different in sequence.

By "naturally occurring fibronectin" is meant any fibronectin protein that is encoded by a living organism.

By "randomized" is meant including one or more amino acid alterations relative to a template sequence.

By a "protein" is meant any sequence of two or more amino acids, regardless of length, post-translation modification, or function. "Protein" and "peptide" are used interchangeably herein.

By "RNA" is meant a sequence of two or more covalently bonded, naturally occurring or modified ribonucleotides. One example of a modified RNA included within this term is phosphorothioate RNA.

By "DNA" is meant a sequence of two or more covalently bonded, naturally occurring or modified deoxyribonucleotides.

By a "nucleic acid" is meant any two or more covalently bonded nucleotides or nucleotide analogs or derivatives. As used herein, this term includes, without limitation, DNA, RNA, and PNA.

By "pharmaceutically acceptable" is meant a compound or protein that may be administered to an animal (for example, a mammal) without significant adverse medical consequences.

By "physiologically acceptable carrier" is meant a carrier which does not have a significant detrimental impact on the treated host and which retains the therapeutic properties of the compound with which it is administered. One exemplary physiologically acceptable carrier is physiological saline. Other physiologically acceptable carriers and their formulations are known to one skilled in the art and are described, for example, in Remington's Pharmaceutical Sciences, (18th edition), ed. A. Gennaro, 1990, Mack Publishing Company, Easton, PA, incorporated herein by reference.

By "selecting" is meant substantially partitioning a molecule from other molecules in a population. As used herein, a "selecting" step provides at least a 2-fold, preferably, a 30-fold, more preferably, a 100-fold, and, most preferably, a 1000-fold enrichment of a desired molecule relative to undesired molecules in a population following the selection step. A selection step may be repeated any number of times, and different types of selection steps may be combined in a given approach.

By "binding partner," as used herein, is meant any molecule which has a specific, covalent or non-covalent affinity for a portion of a desired compound (for example, protein) of interest. Examples of binding partners include, without limitation, members of antigen/antibody pairs,

protein/inhibitor pairs, receptor/ligand pairs (for example cell surface receptor/ligand pairs, such as hormone receptor/peptide hormone pairs), enzyme/substrate pairs (for example, kinase/substrate pairs), lectin/carbohydrate pairs, oligomeric or heterooligomeric protein aggregates, DNA binding protein/DNA binding site pairs, RNA/protein pairs, and nucleic acid duplexes, heteroduplexes, or ligated strands, as well as any molecule which is capable of forming one or more covalent or non-covalent bonds (for example, disulfide bonds) with any portion of another molecule (for example, a compound or protein).

By a "solid support" is meant, without limitation, any column (or column material), bead, test tube, microtiter dish, solid particle (for example, agarose or sepharose), microchip (for example, silicon, silicon-glass, or gold chip), or membrane (for example, the membrane of a liposome or vesicle) to which a fibronectin scaffold or an affinity complex may be bound, either directly or indirectly (for example, through other binding partner intermediates such as other antibodies or Protein A), or in which a fibronectin scaffold or an affinity complex may be embedded (for example, through a receptor or channel).

The present invention provides a number of advantages. For example, as described in more detail below, the present antibody mimics exhibit improved biophysical properties, such as stability under reducing conditions and solubility at high concentrations. In addition, these molecules may be readily expressed and folded in prokaryotic systems, such as E. coli, in eukaryotic systems, such as yeast, and in in vitro translation systems, such as the rabbit reticulocyte lysate system. Moreover, these molecules are extremely amenable to affinity maturation techniques involving multiple cycles of selection, including in vitro selection using RNA-protein fusion technology

(Roberts and Szostak, Proc. Natl. Acad. Sci USA 94:12297, 1997; Szostak et al., U.S.S.N. 09/007,005 and U.S.S.N. 09/247,190; Szostak et al. WO98/31700), phage display (see, for example, Smith and Petrenko, Chem. Rev. 97:317, 1997), and yeast display systems (see, for example, Boder and Wittrup, Nature Biotech. 15:553, 1997).

Other features and advantages of the present invention will be apparent from the following detailed description thereof, and from the claims.

Brief Description of the Drawings

FIGURE 1 is a photograph showing a comparison between the structures of antibody heavy chain variable regions from camel (dark blue) and llama (light blue), in each of two orientations.

FIGURE 2 is a photograph showing a comparison between the structures of the camel antibody heavy chain variable region (dark blue), the llama antibody heavy chain variable region (light blue), and a fibronectin type III module number 10 (^{10}Fn3) (yellow).

FIGURE 3 is a photograph showing a fibronectin type III module number 10 (^{10}Fn3), with the loops corresponding to the antigen-binding loops in IgG heavy chains highlighted in red.

FIGURE 4 is a graph illustrating a sequence alignment between a fibronectin type III protein domain and related protein domains.

FIGURE 5 is a photograph showing the structural similarities between a ^{10}Fn3 domain and 15 related proteins, including fibronectins, tenascins, collagens, and undulin. In this photograph, the regions are labeled as follows: constant, dark blue; conserved, light blue; neutral, white; variable, red; and RGB integrin-binding motif (variable), yellow.

FIGURE 6 is a photograph showing space filling models of fibronectin III modules 9 and 10, in each of two different orientations. The two modules and the integrin binding loop (RGB) are labeled. In this figure, blue indicates positively charged residues, red indicates negatively charged residues, and white indicates uncharged residues.

FIGURE 7 is a photograph showing space filling models of fibronectin III modules 7-10, in each of three different orientiations. The four modules are labeled. In this figure, blue indicates positively charged residues, red indicates negatively charged residues, and white indicates uncharged residues.

FIGURE 8 is a photograph illustrating the formation, under different salt conditions, of RNA-protein fusions which include fibronectin type III domains.

FIGURE 9 is a series of photographs illustrating the selection of fibronectin type III domain-containing RNA-protein fusions, as measured by PCR signal analysis.

FIGURE 10 is a graph illustrating an increase in the percent TNF-α binding during the selections described herein, as well as a comparison between RNA-protein fusion and free protein selections.

FIGURE 11 is a series of schematic representations showing IgG, ^{10}Fn3, Fn-CH_1-CH_2-CH_3, and Fn-CH_2-CH_3 (clockwise from top left).

FIGURE 12 is a photograph showing a molecular model of Fn-CH_1-CH_2-CH_3 based on known three-dimensional structures of IgG (X-ray crystallography) and ^{10}Fn3 (NMR and X-ray crystallography).

FIGURE 13 is a graph showing the time course of an exemplary ^{10}Fn3-based nucleic acid-protein fusion selection of TNF-α binders. The proportion of nucleic acid-protein fusion pool (open diamonds) and free

protein pool (open circles) that bound to TNF-α-Sepharose, and the proportion of free protein pool (full circles) that bound to underivatized Sepharose, are shown.

FIGURES 14 and 15 are graphs illustrating TNF-α binding by TNF-α Fn-binders. In particular, these figures show mass spectra data obtained from a ^{10}Fn3 fusion chip and non-fusion chip, respectively.

FIGURES 16 and 17 are the phosphorimage and fluorescence scan, respectively, of a ^{10}Fn3 array, illustrating TNF-α binding.

Detailed Description

The novel antibody mimics described herein have been designed to be superior both to antibody-derived fragments and to non-antibody frameworks, for example, those frameworks described above.

The major advantage of these antibody mimics over antibody fragments is structural. These scaffolds are derived from whole, stable, and soluble structural modules found in human body fluid proteins. Consequently, they exhibit better folding and thermostability properties than antibody fragments, whose creation involves the removal of parts of the antibody native fold, often exposing amino acid residues that, in an intact antibody, would be buried in a hydrophobic environment, such as an interface between variable and constant domains. Exposure of such hydrophobic residues to solvent increases the likelihood of aggregation.

In addition, the antibody mimics described herein have no disulfide bonds, which have been reported to retard or prevent proper folding of antibody fragments under certain conditions. Since the present scaffolds do not rely on disulfides for native fold stability, they are stable under reducing conditions, unlike antibodies and their fragments which unravel upon disulfide

bond breakdown.

Moreover, these fibronectin-based scaffolds provide the functional advantages of antibody molecules. In particular, despite the fact that the ^{10}Fn3 module is not an immunoglobulin, its overall fold is close to that of the variable region of the IgG heavy chain (Figure 2), making it possible to display the three fibronectin loops analogous to CDRs in relative orientations similar to those of native antibodies. Because of this structure, the present antibody mimics possess antigen binding properties that are similar in nature and affinity to those of antibodies, and a loop randomization and shuffling strategy may be employed in vitro that is similar to the process of affinity maturation of antibodies in vivo.

There are now described below exemplary fibronectin-based scaffolds and their use for identifying, selecting, and evolving novel binding proteins as well as their target ligands. These examples are provided for the purpose of illustrating, and not limiting, the invention.

^{10}Fn3 Structural Motif

The antibody mimics of the present invention are based on the structure of a fibronectin module of type III (Fn3), a common domain found in mammalian blood and structural proteins. This domain occurs more than 400 times in the protein sequence database and has been estimated to occur in 2% of the proteins sequenced to date, including fibronectins, tenscin, intracellular cytoskeletal proteins, and prokaryotic enzymes (Bork and Doolittle, Proc. Natl. Acad. Sci. USA 89:8990, 1992; Bork et al., Nature Biotech. 15:553, 1997; Meinke et al., J. Bacteriol. 175:1910, 1993; Watanabe et al., J. Biol. Chem. 265:15659, 1990). In particular, these scaffolds include, as templates, the tenth module of human Fn3 (^{10}Fn3), which comprises 94 amino acid residues.

The overall fold of this domain is closely related to that of the smallest functional antibody fragment, the variable region of the heavy chain, which comprises the entire antigen recognition unit in camel and llama IgG (Figure 1, 2). The major differences between camel and llama domains and the ^{10}Fn3 domain are that (i) ^{10}Fn3 has fewer beta strands (seven vs. nine) and (ii) the two beta sheets packed against each other are connected by a disulfide bridge in the camel and llama domains, but not in ^{10}Fn3.

The three loops of ^{10}Fn3 corresponding to the antigen-binding loops of the IgG heavy chain run between amino acid residues 21-31, 51-56, and 76-88 (Figure 3). The length of the first and the third loop, 11 and 12 residues, respectively, fall within the range of the corresponding antigen-recognition loops found in antibody heavy chains, that is, 10-12 and 3-25 residues, respectively. Accordingly, once randomized and selected for high antigen affinity, these two loops make contacts with antigens equivalent to the contacts of the corresponding loops in antibodies.

In contrast, the second loop of ^{10}Fn3 is only 6 residues long, whereas the corresponding loop in antibody heavy chains ranges from 16-19 residues. To optimize antigen binding, therefore, the second loop of ^{10}Fn3 is preferably extended by 10-13 residues (in addition to being randomized) to obtain the greatest possible flexibility and affinity in antigen binding. Indeed, in general, the lengths as well as the sequences of the CDR-like loops of the antibody mimics may be randomized during in vitro or in vivo affinity maturation (as described in more detail below).

The tenth human fibronectin type III domain, ^{10}Fn3, refolds rapidly even at low temperature; its backbone conformation has been recovered within 1 second at 5°C. Thermodynamic stability of ^{10}Fn3 is high (ΔG_U = 24 kJ/mol = 5.7 kcal/mol), correlating with its high melting temperature of 110°C.

One of the physiological roles of ^{10}Fn3 is as a subunit of fibronectin, a glycoprotein that exists in a soluble form in body fluids and in an insoluble form in the extracellular matrix (Dickinson et al., J. Mol. Biol. 236:1079, 1994). A fibronectin monomer of 220-250 kD contains 12 type I modules, two type II modules, and 17 fibronectin type III modules (Potts and Campbell, Curr. Opin.Cell Biol. 6:648, 1994). Different type III modules are involved in the binding of fibronectin to integrins, heparin, and chondroitin sulfate. ^{10}Fn3 was found to mediate cell adhesion through an integrin-binding Arg-Gly-Asp (RGD) motif on one of its exposed loops. Similar RGD motifs have been shown to be involved in integrin binding by other proteins, such as fibrinogen, von Wellebrand factor, and vitronectin (Hynes et al., Cell 69:11, 1992). No other matrix- or cell-binding roles have been described for ^{10}Fn3.

The observation that ^{10}Fn3 has only slightly more adhesive activity than a short peptide containing RGD is consistent with the conclusion that the cell-binding activity of ^{10}Fn3 is localized in the RGD peptide rather than distributed throughout the ^{10}Fn3 structure (Baron et al., Biochemistry 31:2068, 1992). The fact that ^{10}Fn3 without the RGD motif is unlikely to bind to other plasma proteins or extracellular matrix makes ^{10}Fn3 a useful scaffold to replace antibodies. In addition, the presence of ^{10}Fn3 in natural fibrinogen in the bloodstream suggests that ^{10}Fn3 itself is unlikely to be immunogenic in the organism of origin.

In addition, we have determined that the ^{10}Fn3 framework possesses exposed loop sequences tolerant of randomization, facilitating the generation of diverse pools of antibody mimics. This determination was made by examining the flexibility of the ^{10}Fn3 sequence. In particular, the human ^{10}Fn3 sequence was aligned with the sequences of fibronectins from other sources as well as sequences of related proteins (Figure 4), and the results of this

alignment were mapped onto the three-dimensional structure of the human ^{10}Fn3 domain (Figure 5). This alignment revealed that the majority of conserved residues are found in the core of the beta sheet sandwich, whereas the highly variable residues are located along the edges of the beta sheets, including the N- and C-termini, on the solvent-accessible faces of both beta sheets, and on three solvent-accessible loops that serve as the hypervariable loops for affinity maturation of the antibody mimics. In view of these results, the randomization of these three loops are unlikely to have an adverse effect on the overall fold or stability of the ^{10}Fn3 framework itself.

For the human ^{10}Fn3 sequence, this analysis indicates that, at a minimum, amino acids 1-9, 44-50, 61-54, 82-94 (edges of beta sheets); 19, 21, 30-46 (even), 79-65 (odd) (solvent-accessible faces of both beta sheets); 21-31, 51-56, 76-88 (CDR-like solvent-accessible loops); and 14-16 and 36-45 (other solvent-accessible loops and beta turns) may be randomized to evolve new or improved compound-binding proteins. In addition, as discussed above, alterations in the lengths of one or more solvent exposed loops may also be included in such directed evolution methods. Alternatively, changes in the β-sheet sequences may also be used to evolve new proteins. These mutations change the scaffold and thereby indirectly alter loop structure(s). If this approach is taken, mutations should not saturate the sequence, but rather few mutations should be introduced. Preferably, no more than 10 amino acid changes, and, more preferably, no more than 3 amino acid changes should be introduced to the β-sheet sequences by this approach.

Fibronectin Fusions

The antibody mimics described herein may be fused to other protein domains. For example, these mimics may be integrated with the human

immune response by fusing the constant region of an IgG (F_c) with a ^{10}Fn3 module, preferably through the C-terminus of ^{10}Fn3. The F_c in such a ^{10}Fn3-F_c fusion molecule activates the complement component of the immune response and increases the therapeutic value of the antibody mimic. Similarly, a fusion between ^{10}Fn3 and a complement protein, such as C1q, may be used to target cells, and a fusion between ^{10}Fn3 and a toxin may be used to specifically destroy cells that carry a particular antigen. In addition, ^{10}Fn3 in any form may be fused with albumin to increase its half-life in the bloodstream and its tissue penetration. Any of these fusions may be generated by standard techniques, for example, by expression of the fusion protein from a recombinant fusion gene constructed using publically available gene sequences.

Fibronectin Scaffold Multimers

In addition to fibronectin monomers, any of the fibronectin constructs described herein may be generated as dimers or multimers of ^{10}Fn3-based antibody mimics as a means to increase the valency and thus the avidity of antigen binding. Such multimers may be generated through covalent binding between individual ^{10}Fn3 modules, for example, by imitating the natural ^{8}Fn3-^{9}Fn3-^{10}Fn3 C-to-N-terminus binding or by imitating antibody dimers that are held together through their constant regions. A ^{10}Fn3-Fc construct may be exploited to design dimers of the general scheme of ^{10}Fn3-Fc::Fc-^{10}Fn3. The bonds engineered into the Fc::Fc interface may be covalent or non-covalent. In addition, dimerizing or multimerizing partners other than Fc can be used in ^{10}Fn3 hybrids to create such higher order structures.

In particular examples, covalently bonded multimers may be generated by constructing fusion genes that encode the multimer or, alternatively, by engineering codons for cysteine residues into monomer sequences and allowing disulfide bond formation to occur between the expression products. Non-covalently bonded multimers may also be generated by a variety of techniques. These include the introduction, into monomer sequences, of codons corresponding to positively and/or negatively charged residues and allowing interactions between these residues in the expression products (and therefore between the monomers) to occur. This approach may be simplified by taking advantage of charged residues naturally present in a monomer subunit, for example, the negatively charged residues of fibronectin. Another means for generating non-covalently bonded antibody mimics is to introduce, into the monomer gene (for example, at the amino- or carboxy-termini), the coding sequences for proteins or protein domains known to interact. Such proteins or protein domains include coil-coil motifs, leucine zipper motifs, and any of the numerous protein subunits (or fragments thereof) known to direct formation of dimers or higher order multimers.

Fibronectin-Like Molecules

Although ^{10}Fn3 represents a preferred scaffold for the generation of antibody mimics, other molecules may be substituted for ^{10}Fn3 in the molecules described herein. These include, without limitation, human fibronectin modules ^{1}Fn3-^{9}Fn3 and ^{11}Fn3-^{17}Fn3 as well as related Fn3 modules from non-human animals and prokaryotes. In addition, Fn3 modules from other proteins with sequence homology to ^{10}Fn3, such as tenascins and undulins, may also be used. Modules from different organisms and parent proteins may be most appropriate for different applications; for example, in

designing an antibody mimic, it may be most desirable to generate that protein from a fibronectin or fibronectin-like molecule native to the organism for which a therapeutic or diagnostic molecule is intended.

Directed Evolution of Scaffold-Based Binding Proteins

The antibody mimics described herein may be used in any technique for evolving new or improved binding proteins. In one particular example, the target of binding is immobilized on a solid support, such as a column resin or microtiter plate well, and the target contacted with a library of candidate scaffold-based binding proteins. Such a library may consist of ^{10}Fn3 clones constructed from the wild type ^{10}Fn3 scaffold through randomization of the sequence and/or the length of the ^{10}Fn3 CDR-like loops. If desired, this library may be an RNA-protein fusion library generated, for example, by the techniques described in Szostak et al., U.S.S.N. 09/007,005 and 09/247,190; Szostak et al., WO98/31700; and Roberts & Szostak, Proc. Natl. Acad. Sci. USA (1997) vol. 94, p. 12297-12302. Alternatively, it may be a DNA-protein library (for example, as described in Lohse, DNA-Protein Fusions and Uses Thereof, U.S.S.N. 60/110,549, U.S.S.N. 09/459,190, and US 99/28472). The fusion library is incubated with the immobilized target, the support is washed to remove non-specific binders, and the tightest binders are eluted under very stringent conditions and subjected to PCR to recover the sequence information or to create a new library of binders which may be used to repeat the selection process, with or without further mutagenesis of the sequence. A number of rounds of selection may be performed until binders of sufficient affinity for the antigen are obtained.

In one particular example, the ^{10}Fn3 scaffold may be used as the selection target. For example, if a protein is required that binds a specific peptide sequence presented in a ten residue loop, a single ^{10}Fn3 clone is constructed in which one of its loops has been set to the length of ten and to the desired sequence. The new clone is expressed *in vivo* and purified, and then immobilized on a solid support. An RNA-protein fusion library based on an appropriate scaffold is then allowed to interact with the support, which is then washed, and desired molecules eluted and re-selected as described above.

Similarly, the ^{10}Fn3 scaffold may be used to find natural proteins that interact with the peptide sequence displayed in a ^{10}Fn3 loop. The ^{10}Fn3 protein is immobilized as described above, and an RNA-protein fusion library is screened for binders to the displayed loop. The binders are enriched through multiple rounds of selection and identified by DNA sequencing.

In addition, in the above approaches, although RNA-protein libraries represent exemplary libraries for directed evolution, any type of scaffold-based library may be used in the selection methods of the invention.

Use

The antibody mimics described herein may be evolved to bind any antigen of interest. These proteins have thermodynamic properties superior to those of natural antibodies and can be evolved rapidly in vitro. Accordingly, these antibody mimics may be employed in place of antibodies in all areas in which antibodies are used, including in the research, therapeutic, and diagnostic fields. In addition, because these scaffolds possess solubility and stability properties superior to antibodies, the antibody mimics described herein may also be used under conditions which would destroy or inactivate antibody molecules. Finally, because the scaffolds of the present invention

may be evolved to bind virtually any compound, these molecules provide completely novel binding proteins which also find use in the research, diagnostic, and therapeutic areas.

Experimental Results

Exemplary scaffold molecules described above were generated and tested, for example, in selection protocols, as follows.

Library construction

A complex library was constructed from three fragments, each of which contained one randomized area corresponding to a CDR-like loop. The fragments were named BC, DE, and FG, based on the names of the CDR-H-like loops contained within them; in addition to ^{10}Fn3 and a randomized sequence, each of the fragments contained stretches encoding an N-terminal His_6 domain or a C-terminal FLAG peptide tag. At each junction between two fragments (i.e., between the BC and DE fragments or between the DE and FG fragments), each DNA fragment contained recognition sequences for the EarI Type IIS restriction endonuclease. This restriction enzyme allowed the splicing together of adjacent fragments while removing all foreign, non-^{10}Fn3, sequences. It also allows for a recombination-like mixing of the three ^{10}Fn3 fragments between cycles of mutagenesis and selection.

Each fragment was assembled from two overlapping oligonucleotides, which were first annealed, then extended to form the double-stranded DNA form of the fragment. The oligonucleotides that were used to construct and process the three fragments are listed below; the "Top" and "Bottom" species for each fragment are the oligonucleotides that contained the entire ^{10}Fn3 encoding sequence. In these oligonucleotides designations,

"N" indicates A, T, C, or G; and "S" indicates C or G.

HfnLbcTop (His):
5'- GG AAT TCC TAA TAC GAC TCA CTA TAG GGA CAA TTA CTA TTT ACA ATT ACA ATG CAT CAC CAT CAC CAT CAC GTT TCT GAT GTT CCG AGG GAC CTG GAA GTT GTT GCT GCG ACC CCC ACC AGC-3' (SEQ ID NO: 1)

HfnLbcTop (an alternative N-terminus):
5'- GG AAT TCC TAA TAC GAC TCA CTA TAG GGA CAA TTA CTA TTT ACA ATT ACA ATG GTT TCT GAT GTT CCG AGG GAC CTG GAA GTT GTT GCT GCG ACC CCC ACC AGC-3' (SEQ ID NO: 2)

HFnLBCBot-flag8:
5'-AGC GGA TGC CTT GTC GTC GTC GTC CTT GTA GTC GCT CTT CCC TGT TTC TCC GTA AGT GAT CCT GTA ATA TCT (SNN)7 CCA GCT GAT CAG TAG GCT GGT GGG GGT CGC AGC -3' (SEQ ID NO: 3)

HFnBC3'-flag8:
5'-AGC GGA TGC CTT GTC GTC GTC GTC CTT GTA GTC GCT CTT CCC TGT TTC TCC GTA AGT GAT CC-3' (SEQ ID NO: 4)

HFnLDETop:
5'- GG AAT TCC TAA TAC GAC TCA CTA TAG GGA CAA TTA CTA TTT ACA ATT ACA ATG CAT CAC CAT CAC CAT CAC CTC TTC ACA GGA GGA AAT AGC CCT GTC C-3' (SEQ ID NO: 5)

HFnLDEBot-flag8:
5'-AGC GGA TGC CTT GTC GTC GTC GTC CTT GTA GTC GCT CTT CGT ATA ATC AAC TCC AGG TTT AAG GCC GCT GAT GGT AGC TGT (SNN)4 AGG CAC AGT GAA CTC CTG GAC AGG GCT ATT TCC TCC TGT -3' (SEQ ID NO: 6)

HFnDE3'-flag8:
5'-AGC GGA TGC CTT GTC GTC GTC GTC CTT GTA GTC GCT CTT CGT ATA ATC AAC TCC AGG TTT AAG G-3' (SEQ ID NO: 7)

HFnLFGTop:
5'- GG AAT TCC TAA TAC GAC TCA CTA TAG GGA CAA TTA CTA TTT ACA ATT ACA ATG CAT CAC CAT CAC CAT CAC CTC TTC TAT ACC ATC ACT GTG TAT GCT GTC-3' (SEQ ID NO: 8)

HFnLFGBot-flag8:
5'-AGC GGA TGC CTT GTC GTC GTC GTC CTT GTA GTC TGT TCG GTA ATT AAT GGA AAT TGG (SNN)10 AGT GAC AGC ATA CAC AGT GAT GGT ATA -3' (SEQ ID NO: 9)

HFnFG3'-flag8:
5'-AGC GGA TGC CTT GTC GTC GTC GTC CTT GTA GTC TGT TCG GTA ATT AAT GGA AAT TGG -3' (SEQ ID NO: 10)

T7Tmv (introduces T7 promoter and TMV untranslated region needed for in vitro translation):
5'- GCG TAA TAC GAC TCA CTA TAG GGA CAA TTA CTA TTT ACA

ATT ACA-3' (SEQ ID NO: 11)

ASAflag8:
5'-AGC GGA TGC CTT GTC GTC GTC GTC CTT GTA GTC-3' (SEQ ID NO: 12)

Unispl-s (spint oligonucleotide used to ligate mRNA to the puromycin-containing linker, described by Roberts et al, 1997, supra):
5'-TTTTTTTTTNAGCGGATGC-3' (SEQ ID NO: 13)

A18---2PEG (DNA-puromycin linker):
5'-(A)18(PEG)2CCPur (SEQ ID NO: 14)

The pairs of oligonucleotides (500 pmol of each) were annealed in 100 μL of 10 mM Tris 7.5, 50 mM NaCl for 10 minutes at 85°C, followed by a slow (0.5-1 hour) cooling to room temperature. The annealed fragments with single-stranded overhangs were then extended using 100 U Klenow (New England Biolabs, Beverly, MA) for each 100 μL aliquot of annealed oligos, and the buffer made of 838.5 μl H_2O, 9 μl 1 M Tris 7.5, 5 μl 1M $MgCl_2$, 20 μl 10 mM dNTPs, and 7.5 μl 1M DTT. The extension reactions proceeded for 1 hour at 25°C.

Next, each of the double-stranded fragments was transformed into a RNA-protein fusion (PROfusion™) using the technique developed by Szostak et al., U.S.S.N. 09/007,005 and U.S.S.N. 09/247,190; Szostak et al., WO98/31700; and Roberts & Szostak, Proc. Natl. Acad. Sci. USA (1997) vol. 94, p. 12297-12302. Briefly, the fragments were transcribed using an Ambion in vitro transcription kit, MEGAshortscript (Ambion, Austin, TX), and the

resulting mRNA was gel-purified and ligated to a DNA-puromycin linker using DNA ligase. The mRNA-DNA-puromycin molecule was then translated using the Ambion rabbit reticulocyte lysate-based translation kit. The resulting mRNA-DNA-puromycin-protein PROfusion™ was purified using Oligo(dT) cellulose, and a complementary DNA strand was synthesized using reverse transcriptase and the RT primers described above (Unisplint-S or flagASA), following the manufacturer's instructions.

The PROfusion™ obtained for each fragment was next purified on the resin appropriate to its peptide purification tag, i.e., on Ni-NTA agarose for the His_6-tag and M2 agarose for the FLAG-tag, following the procedure recommended by the manufacturer. The DNA component of the tag-binding PROfusions™ was amplified by PCR using Pharmacia Ready-to-Go PCR Beads, 10 pmol of 5' and 3' PCR primers, and the following PCR program (Pharmacia, Piscataway, NJ): Step 1: 95°C for 3 minutes; Step 2: 95°C for 30 seconds, 58/62°C for 30 seconds, 72°C for 1 minute, 20/25/30 cycles, as required; Step 3: 72°C for 5 minutes; Step 4: 4°C until end.

The resulting DNA was cleaved by 5 U EarI (New England Biolabs) per1 ug DNA; the reaction took place in T4 DNA Ligase Buffer (New England Biolabs) at 37°C, for 1 hour, and was followed by an incubation at 70°C for15 minutes to inactivate Ear I. Equal amounts of the BC, DE, and FG fragments were combined and ligated to form a full-length ^{10}Fn3 gene with randomized loops. The ligation required 10 U of fresh EarI (New England Biolabs) and 20 U of T4 DNA Ligase (Promega, Madison, WI), and took 1 hour at 37°C.

Three different libraries were made in the manner described above. Each contained the form of the FG loop with 10 randomized residues. The BC and the DE loops of the first library bore the wild type ^{10}Fn3 sequence; a BC loop with 7 randomized residues and a wild type DE loop made up the second

library; and a BC loop with 7 randomized residues and a DE loop with 4 randomized residues made up the third library. The complexity of the FG loop in each of these three libraries was 10^{13}; the further two randomized loops provided the potential for a complexity too large to be sampled in a laboratory.

The three libraries constructed were combined into one master library in order to simplify the selection process; target binding itself was expected to select the most suitable library for a particular challenge. PROfusions™ were obtained from the master library following the general procedure described in Szostak et al., U.S.S.N. 09/007,005 and 09/247,190; Szostak et al., WO98/31700; and Roberts & Szostak, Proc. Natl. Acad. Sci. USA (1997) vol. 94, p. 12297-12302 (Figure 8).

Fusion Selections

The master library in the PROfusion™ form was subjected to selection for binding to TNF-α. Two protocols were employed: one in which the target was immobilized on an agarose column and one in which the target was immobilized on a BIACORE chip. First, an extensive optimization of conditions to minimize background binders to the agarose column yielded the favorable buffer conditions of 50 mM HEPES pH 7.4, 0.02% Triton, 100 μg/ml Sheared Salmon Sperm DNA. In this buffer, the non-specific binding of the ^{10}Fn3 RNA fusion to TNF-α Sepharose was 0.3%. The non-specific binding background of the ^{10}Fn3 RNA-DNA to TNF-α Sepharose was found to be 0.1%.

During each round of selection on TNF-α Sepharose, the Profusion™ library was first preincubated for an hour with underivatized Sepharose to remove any remaining non-specific binders; the flow-through from this pre-clearing was incubated for another hour with TNF-α Sepharose.

The TNF-α Sepharose was washed for 3-30 minutes.

After each selection, the PROfusion™ DNA that had been eluted from the solid support with 0.3 M NaOH or 0.1M KOH was amplified by PCR; a DNA band of the expected size persisted through multiple rounds of selection (Figure 9); similar results were observed in the two alternative selection protocols, and only the data from the agarose column selection is shown in Figure 9.

In the first seven rounds, the binding of library PROfusions™ to the target remained low; in contrast, when free protein was translated from DNA pools at different stages of the selection, the proportion of the column binding species increased significantly between rounds (Figure 10). Similar selections may be carried out with any other binding species target (for example, IL-1 and IL-13).

Animal Studies

Wild-type ^{10}Fn3 contains an integrin-binding tripepetide motif, Arginine 78 - Glycine 79 - Aspartate 80 (the "RGD motif') at the tip of the FG loop. In order to avoid integrin binding and a potential inflammatory response based on this tripeptide in vivo, a mutant form of ^{10}Fn3 was generated that contained an inert sequence, Serine 78 - Glycine 79 - Glutamate 80 (the "SGE mutant"), a sequence which is found in the closely related, wild-type ^{11}Fn3 domain. This SGE mutant was expressed as an N-terminally His_6-tagged, free protein in E. coli, and purified to homogeneity on a metal chelate column followed by a size exclusion column.

In particular, the DNA sequence encoding His_6-^{10}Fn3(SGE) was cloned into the pET9a expression vector and transformed into BL21 DE3 pLysS cells. The culture was then grown in LB broth containing 50 μg/mL

kanamycin at 37°C, with shaking, to A_{560}=1.0, and was then induced with 0.4 mM IPTG. The induced culture was further incubated, under the same conditions, overnight (14-18 hours); the bacteria were recovered by standard, low speed centrifugation. The cell pellet was resuspended in 1/50 of the original culture volume of lysis buffer (50 mM Tris 8.0, 0.5 M NaCl, 5% glycerol, 0.05% Triton X-100, and 1 mM PMSF), and the cells were lysed by passing the resulting paste through a Microfluidics Corporation Microfluidizer M110-EH, three times. The lysate was clarified by centrifugation, and the supernatant was filtered through a 0.45 μm filter followed by filtration through a 0.2 μm filter. 100 mL of the clarified lysate was loaded onto a 5 mL Talon cobalt column (Clontech, Palo Alto, CA), washed by 70 mL of lysis buffer, and eluted with a linear gradient of 0-30 mM imidazole in lysis buffer. The flow rate through the column through all the steps was 1 mL/min. The eluted protein was concentrated 10-fold by dialysis (MW cutoff = 3,500) against 15,000-20,000 PEG. The resulting sample was dialysed into buffer 1 (lysis buffer without the glycerol), then loaded, 5 mL at a time, onto a 16 x 60 mm Sephacryl 100 size exclusion column equilibrated in buffer 1. The column was run at 0.8 mL/min, in buffer 1; all fractions that contained a protein of the expected MW were pooled, concentrated 10X as described above, then dialyzed into PBS. Toxikon (MA) was engaged to perform endotoxin screens and animal studies on the resulting sample.

In these animal studies, the endotoxin levels in the samples examined to date have been below the detection level of the assay. In a preliminary toxicology study, this protein was injected into two mice at the estimated 100X therapeutic dose of 2.6 mg/mouse. The animals survived the two weeks of the study with no apparent ill effects. These results suggest that ^{10}Fn3 may be incorporated safely into an IV drug.

Alternative Constructs for In Vivo Use

To extend the half life of the 8 kD ^{10}Fn3 domain, a larger molecule has also been constructed that mimics natural antibodies. This ^{10}Fn3-F_c molecule contains the -CH_1-CH_2-CH_3 (Figure 11) or -CH_2-CH_3 domains of the IgG constant region of the host; in these constructs, the ^{10}Fn3 domain is grafted onto the N-terminus in place of the IgG V_H domain (Figures 11 and 12). Such antibody-like constructs are expected to improve the pharmacokinetics of the protein as well as its ability to harness the natural immune response.

In order to construct the murine form of the ^{10}Fn3-CH_1-CH_2-CH_3 clone, the -CH_1-CH_2-CH_3 region was first amplified from a mouse liver spleen cDNA library (Clontech), then ligated into the pET25b vector. The primers used in the cloning were 5' Fc Nest and 3' 5 Fc Nest, and the primers used to graft the appropriate restriction sites onto the ends of the recovered insert were 5' Fc HIII and 3' Fc Nhe:

5' Fc Nest 5'GCG GCA GGG TTT GCT TAC TGG GGC CAA GGG 3' (SEQ ID NO: 15);

3' Fc Nest 5'GGG AGG GGT GGA GGT AGG TCA CAG TCC 3' (SEQ ID NO: 16);

3' Fc Nhe 5' TTT GCT AGC TTT ACC AGG AGA GTG GGA GGC 3' (SEQ ID NO: 17); and

5' Fc HIII 5' AAA AAG CTT GCC AAA ACG ACA CCC CCA TCT GTC 3' (SEQ ID NO: 18).

Further PCR is used to remove the CH_1 region from this clone and create the Fc part of the shorter, ^{10}Fn3-CH_2-CH_3 clone. The sequence encoding ^{10}Fn3 is spliced onto the 5' end of each clone; either the wild type

^{10}Fn3 cloned from the same mouse spleen cDNA library or a modified ^{10}Fn3 obtained by mutagenesis or randomization of the molecules can be used. The oligonucleotides used in the cloning of murine wild-type ^{10}Fn3 were:

Mo 5PCR-NdeI:

5' CATATGGTTTCTGATATTCCGAGAGATCTGGAG 3' (SEQ ID NO: 19);

Mo5PCR-His-NdeI (for an alternative N-terminus with the His_6 purification tag):

5' CAT ATG CAT CAC CAT CAC CAT CAC GTT TCT GAT ATT CCG AGA G 3' (SEQ ID NO: 20); and

Mo3PCR-EcoRI: 5' GAATTCCTATGTTTTATAATTGATGGAAAC3' (SEQ ID NO: 21).

The human equivalents of the clones are constructed using the same strategy with human oligonucleotide sequences.

^{10}Fn3 Scaffolds in Protein Chip Applications

The suitability of the ^{10}Fn3 scaffold for protein chip applications is the consequence of (1) its ability to support many binding functions which can be selected rapidly on the bench or in an automated setup, and (2) its superior biophysical properties.

The versatile binding properties of ^{10}Fn3 are a function of the loops displayed by the Fn3 immunoglobulin-like, beta sandwich fold. As discussed above, these loops are similar to the complementarity determining regions of antibody variable domains and can cooperate in a way similar to those antibody loops in order to bind antigens. In our system, ^{10}Fn3 loops BC (residues

21-30), DE (residues 51-56), and FG (residues 76-87) are randomized either in sequence, in length, or in both sequence and length in order to generate diverse libraries of mRNA-^{10}Fn3 fusions. The binders in such libraries are then enriched based on their affinity for an immobilized or tagged target, until a small population of high affinity binders are generated. Also, error-prone PCR and recombination can be employed to facilitate affinity maturation of selected binders. Due to the rapid and efficient selection and affinity maturation protocols, binders to a large number of targets can be selected in a short time.

As a scaffold for binders to be immobilized on protein chips, the ^{10}Fn3 domain has the advantage over antibody fragments and single-chain antibodies of being smaller and easier to handle. For example, unlike single-chain scaffolds or isolated variable domains of antibodies, which vary widely in their stability and solubility, and which require an oxidizing environment to preserve their structurally essential disulfide bonds, ^{10}Fn3 is extremely stable, with a melting temperature of 110°C, and solubility at a concentration > 16 mg/mL. The ^{10}Fn3 scaffold also contains no disulfides or free cysteines; consequently, it is insensitive to the redox potential of its environment. A further advantage of ^{10}Fn3 is that its antigen-binding loops and N-terminus are on the edge of the beta-sandwich opposite to the C-terminus; thus the attachment of a ^{10}Fn3 scaffold to a chip by its C-terminus aligns the antigen-binding loops, allowing for their greatest accessibility to the solution being assayed. Since ^{10}Fn3 is a single domain of only 94 amino acid residues, it is also possible to immobilize it onto a chip surface at a higher density than is used for single-chain antibodies, with their approximately 250 residues. In addition, the hydrophilicity of the ^{10}Fn3 scaffold, which is reflected in the high solubility of this domain, leads to a lower than average background binding of ^{10}Fn3 to a chip surface.

The stability of the ^{10}Fn3 scaffold as well as its suitability for library formation and selection of binders are likely to be shared by the large, Fn3-like class of protein domains with an immunoglobulin-like fold, such as the domains of tenascin, N-cadherin, E-cadherin, ICAM, titin, GCSF-R, cytokine receptor, glycosidase inhibitor, and antibiotic chromoprotein. The key features shared by all such domains are a stable framework provided by two beta-sheets, which are packed against each other and which are connected by at least three solvent-accessible loops per edge of the sheet; such loops can be randomized to generate a library of potential binders without disrupting the structure of the framework (as described above).

Immobilization of Fibronectin Scaffold Binders (Fn-binders)

To immobilize Fn-binders to a chip surface, a number of exemplary techniques may be utilized. For example, Fn-binders may be immobilized as RNA-protein fusions by Watson-Crick hybridization of the RNA moiety of the fusion to a base complementary DNA immobilized on the chip surface (as described, for example, in Addressable Protein Arrays, U.S.S.N. 60/080,686; U.S.S.N. 09/282,734; and WO 99/51773). Alternatively, Fn-binders can be immobilized as free proteins directly on a chip surface. Manual as well as robotic devices may be used for deposition of the Fn-binders on the chip surface. Spotting robots can be used for deposition of Fn-binders with high density in an array format (for example, by the method of Lueking et al., Anal Biochem. 1999 May 15;270(1):103-11). Different methods may also be utilized for anchoring the Fn-binder on the chip surface. A number of standard immobilization procedures may be used including those described in Methods in Enzymology (K. Mosbach and B. Danielsson, eds.), vols. 135 and 136, Academic Press, Orlando, Florida, 1987; Nilsson et al., Protein Expr. Purif.

1997 Oct;11(1):1-16; and references therein. Oriented immobilization of Fn-binders can help to increase the binding capacity of chip-bound Fn-binders. Exemplary approaches for achieving oriented coupling are described in Lu et al., The Analyst (1996), vol. 121, p. 29R-32R; and Turkova, J Chromatogr B Biomed Sci App. 1999 Feb 5;722(1-2):11-31. In addition, any of the methods described herein for anchoring Fn-binders to chip surfaces can also be applied to the immobilization of Fn-binders on beads, or other supports.

Target Protein Capture and Detection

Selected populations of Fn-binders may be used for detection and/or quantitation of analyte targets, for example, in samples such as biological samples. To carry out this type of diagnostic assay, selected Fn-binders to targets of interest are immobilized on an appropriate support to form multi-featured protein chips. Next, a sample is applied to the chip, and the components of the sample that associate with the Fn-binders are identified based on the target-specificity of the immobilized binders. Using this technique, one or more components may be simultaneously identified or quantitated in a sample (for example, as a means to carry out sample profiling).

Methods for target detection allow measuring the levels of bound protein targets and include, without limitation, radiography, fluorescence scanning, mass spectroscopy (MS), and surface plasmon resonance (SPR). Autoradiography using a phosphorimager system (Molecular Dynamics, Sunnyvale, CA) can be used for detection and quantification of target protein which has been radioactively labeled, e.g., using ^{35}S methionine. Fluorescence scanning using a laser scanner (see below) may be used for detection and quantification of fluorescently labeled targets. Alternatively, fluorescence scanning may be used for the detection of fluorescently labeled ligands which

themselves bind to the target protein (e.g., fluorescently labeled target-specific antibodies or fluorescently labeled streptavidin binding to target-biotin, as described below).

Mass spectroscopy can be used to detect and identify bound targets based on their molecular mass. Desorption of bound target protein can be achieved with laser assistance directly from the chip surface as described below. Mass detection also allows determinations, based on molecular mass, of target modifications including post-translational modifications like phosophorylation or glycosylation. Surface plasmon resonance can be used for quantification of bound protein targets where the Fn-binder(s) are immobilized on a suitable gold-surface (for example, as obtained from Biacore, Sweden).

Described below are exemplary schemes for selecting Fn binders (in this case, Fn-binders specific for the protein, TNF-α) and the use of those selected populations for detection on chips. This example is provided for the purpose of illustrating the invention, and should not be construed as limiting.

Selection of TNF-α Binders Based on ^{10}Fn3 Scaffold

In one exemplary use for fibronectin scaffold selection on chips, an ^{10}Fn3-based selection was performed against TNF-α, using a library of human ^{10}Fn3 variants with randomized loops BC, DE, and FG. The library was constructed from three DNA fragments, each of which contained nucleotide sequences that encoded approximately one third of human ^{10}Fn3, including one of the randomized loops. The DNA sequences that encoded the loop residues listed above were rebuilt by oligonucleotide synthesis, so that the codons for the residues of interest were replaced by (NNS)n, where N represents any of the four deoxyribonucleotides (A, C, G, or T), and S represents either C or G. The C-terminus of each fragment contained the sequence for the FLAG

purification tag.

Once extended by Klenow, each DNA fragment was transcribed, ligated to a puromycin-containing DNA linker, and translated in vitro, as described by Szostak et al. (Roberts and Szostak, Proc. Natl. Acad. Sci USA 94:12297, 1997; Szostak et al., U.S.S.N. 09/007,005 and U.S.S.N. 09/247,190; Szostak et al., WO98/31700), to generate an mRNA-peptide fusion, which was then reverse-transcribed into a DNA-mRNA-peptide fusion. The binding of the FLAG-tagged peptide to M2 agarose separated full-length fusion molecules from those containing frameshifts or superfluous stop codons; the DNA associated with the purified full-length fusion was amplified by PCR, then the three DNA fragments were cut by Ear I restriction endonuclease and ligated to form the full length template. The template was transcribed, ligated to puromycin-containing DNA linkers, and translated to generate a ^{10}Fn3-PROfusion™ library, which was then reverse-transcribed to yield the DNA-mRNA-peptide fusion library which was subsequently used in the selection.

Selection for TNF-α binders took place in 50 mM HEPES, pH 7.4, 0.02% Triton-X, 0.1 mg/mL salmon sperm DNA. The PROfusion™ library was incubated with Sepharose-immobilized TNF-α; after washing, the DNA associated with the tightest binders was eluted with 0.1 M KOH, amplified by PCR, and transcribed, ligated, translated, and reverse-transcribed into the starting material for the next round of selection.

Ten rounds of such selection were performed (as shown in Figure 13); they resulted in a PROfusion™ pool that bound to TNF-α-Sepharose with the apparent average Kd of 120 nM. Specific clonal components of the pool that were characterized showed TNF-α binding in the range of 50-500 nM.

Fn-binder Immobilization, Target Protein Capture, and MALDI-TOF Detection

As a first step toward immobilizing the Fn-binders to a chip surface, an oligonucleotide capture probe was prepared with an automated DNA synthesizer (PE BioSystems Expedite 8909) using the solid-support phosphoramidite approach. All reagents were obtained from Glen Research. Synthesis was initiated with a solid support containing a disulfide bond to eventually provide a 3'-terminal thiol functionality. The first four monomers to be added were hexaethylene oxide units, followed by 20 T monomers. The 5'-terminal DMT group was not removed. The capture probe was cleaved from the solid support and deprotected with ammonium hydroxide, concentrated to dryness in a vacuum centrifuge, and purified by reverse-phase HPLC using an acetonitrile gradient in triethylammonium acetate buffer. Appropriate fractions from the HPLC were collected, evaporated to dryness in a vacuum centrifuge, and the 5'-terminal DMT group was removed by treatment with 80% AcOH for 30 minutes. The acid was removed by evaporation, and the oligonucleotide was then treated with 100 mM DTT for 30 minutes to cleave the disulfide bond. DTT was removed by repeated extraction with EtOAc. The oligonucleotide was ethanol precipitated from the remaining aqueous layer and checked for purity by reverse-phase HPLC.

The 3'-thiol capture probe was adjusted to 250 μM in degassed 1X PBS buffer and applied as a single droplet (75 μL) to a 9x9mm gold-coated chip (Biacore) in an argon-flushed chamber containing a small amount of water. After 18 hours at room temperature, the capture probe solution was removed, and the functionalized chip was washed with 50 mL 1X PBS buffer (2x for 15 minutes each) with gentle agitation, and then rinsed with 50 mL water (2x for 15 minutes each) in the same fashion. Remaining liquid was

carefully removed and the functionalized chips were either used immediately or stored at 4°C under argon.

About 1pmol of ^{10}Fn3 fusion pool from the Round 10 TNF-α selection (above) was treated with RNAse A for several hours, adjusted to 5X SSC in 70 μL, and applied to a functionalized gold chip from above as a single droplet. A 50 μL volume gasket device was used to seal the fusion mixture with the functionalized chip, and the apparatus was continuously rotated at 4°C. After 18 hours the apparatus was disassembled, and the gold chip was washed with 50 mL 5X SSC for 10 minutes with gentle agitation. Excess liquid was carefully removed from the chip surface, and the chip was passivated with a blocking solution (1X TBS + 0.02% Tween-20 + 0.25% BSA) for 10 minutes at 4°C. Excess liquid was carefully removed, and a solution containing 500 μg/mL TNF-α in the same composition blocking solution was applied to the chip as a single droplet and incubated at 4°C for two hours with occasional mixing of the droplet via Pipetman. After removal of the binding solution, the chip was washed for 5 minutes at 4°C with gentle agitation (50 mL 1X TBS + 0.02% Tween-20) and then dried at room temperature. A second chip was prepared exactly as described above, except fusion was not added to the hybridization mix.

Next, MALDI-TOF matrix (15 mg/mL 3,5-dimethoxy-4-hydroxycinnamic acid in 1:1 ethanol/10% formic acid in water) was uniformly applied to the gold chips with a high-precision 3-axis robot (MicroGrid, BioRobotics). A 16-pin tool was used to transfer the matrix from a 384-well microtiter plate to the chips, producing 200 micron diameter features with a 600 micron pitch. The MALDI-TOF mass spectrometer (Voyager DE, PerSeptive Biosystems) instrument settings were as follows: Accelerating Voltage = 25k, Grid Voltage = 92%, Guide Wire Voltage =

0.05%, Delay = 200 on, Laser Power = 2400, Low Mass Gate = 1500, Negative Ions = off. The gold chips were individually placed on a MALDI sample stage modified to keep the level of the chip the same as the level of the stage, thus allowing proper flight distance. The instrument's video monitor and motion control system were used to direct the laser beam to individual matrix features.

Figures 14 and 15 show the mass spectra from the ^{10}Fn3 fusion chip and the non-fusion chip, respectively. In each case, a small number of 200 micron features were analyzed to collect the spectra, but Figure 15 required significantly more acquisitions. The signal at 17.5 kDa corresponds to TNF-α monomer.

Fn-binder Immobilization, Target Protein Capture, and Fluorescence Detection

Pre-cleaned 1x3 inch glass microscope slides (Goldseal, #3010) were treated with Nanostrip (Cyantek) for 15 minutes, 10% aqueous NaOH at 70°C for 3 minutes, and 1% aqueous HCl for 1 minute, thoroughly rinsing with deionized water after each reagent. The slides were then dried in a vacuum desiccator over anhydrous calcium sulfate for several hours. A 1% solution of aminopropytrimethoxysilane in 95% acetone / 5% water was prepared and allowed to hydrolyze for 20 minutes. The glass slides were immersed in the hydrolyzed silane solution for 5 minutes with gentle agitation. Excess silane was removed by subjecting the slides to ten 5-minute washes, using fresh portions of 95% acetone / 5% water for each wash, with gentle agitation. The slides were then cured by heating at 110°C for 20 minutes. The silane treated slides were immersed in a freshly prepared 0.2% solution of phenylene 1,4-diisothiocyanate in 90% DMF / 10% pyridine for two hours, with gentle agitation. The slides were washed sequentially with 90% DMF /

10% pyridine, methanol, and acetone. After air drying, the functionalized slides were stored at 0°C in a vacuum desiccator over anhydrous calcium sulfate. Similar results were obtained with commercial amine-reactive slides (3-D Link, Surmodics).

Oligonucleotide capture probes were prepared with an automated DNA synthesizer (PE BioSystems Expedite 8909) using conventional phosphoramidite chemistry. All reagents were from Glen Research. Synthesis was initiated with a solid support bearing an orthogonally protected amino functionality, whereby the 3'-terminal amine is not unmasked until final deprotection step. The first four monomers to be added were hexaethylene oxide units, followed by the standard A, G, C and T monomers. All capture oligo sequences were cleaved from the solid support and deprotected with ammonium hydroxide, concentrated to dyrness, precipitated in ethanol, and purified by reverse-phase HPLC using an acetonitrile gradient in triethylammonium acetate buffer. Appropriate fractions from the HPLC were collected, evaporated to dryness in a vacuum centrifuge, and then coevaporated with a portion of water.

The purified, amine-labeled capture oligos were adjusted to a concentration of 250 μM in 50 mM sodium carbonate buffer (pH 9.0) containing 10% glycerol. The probes were spotted onto the amine-reactive glass surface at defined positions in a 5x5x6 array pattern with a 3-axis robot (MicroGrid, BioRobotics). A 16-pin tool was used to transfer the liquid from 384-well microtiter plates, producing 200 micron features with a 600 micron pitch. Each sub-grid of 24 features represents a single capture probe (i.e., 24 duplicate spots). The arrays were incubated at room temperature in a moisture-saturated environment for 12-18 hours. The attachment reaction was terminated by immersing the chips in 2% aqueous ammonium hydroxide for

five minutes with gentle agitation, followed by rinsing with distilled water (3X for 5 minutes each). The array was finally soaked in 10X PBS solution for 30 minutes at room temperature, and then rinsed again for 5 minutes in distilled water.

Specific and thermodynamically isoenergetic sequences along the ^{10}Fn3 mRNA were identified to serve as capture points to self-assemble and anchor the ^{10}Fn3 protein. The software program HybSimulator v4.0 (Advanced Gene Computing Technology, Inc.) facilitated the identification and analysis of potential capture probes. Six unique capture probes were chosen and printed onto the chip, three of which are complementary to common regions of the ^{10}Fn3 fusion pool's mRNA (CP3', CP5', and CPflag). The remaining three sequences (CPneg1, CPneg2, and CPneg3) are not complementary and function in part as negative controls. Each of the capture probes possesses a 3'-amino terminus and four hexaethylene oxide spacer units, as described above. The following is a list of the capture probe sequences that were employed (5'→3'):

CP3': TGTAAATAGTAATTGTCCC (SEQ ID NO: 22)
CP5': TTTTTTTTTTTTTTTTTTTT (SEQ ID NO: 23)
CPneg1: CCTGTAGGTGTCCAT (SEQ ID NO: 24)
CPflag: CATCGTCCTTGTAGTC (SEQ ID NO: 25)
CPneg2: CGTCGTAGGGGTA (SEQ ID NO: 26)
CPneg3: CAGGTCTTCTTCAGAGA (SEQ ID NO: 27)

About 1pmol of ^{10}Fn3 fusion pool from the Round 10 TNF-α selection was adjusted to 5X SSC containing 0.02% Tween-20 and 2 mM vanadyl ribonucleotide complex in a total volume of 350 μL. The entire volume was

applied to the microarray under a 400 μL gasket device and the assembly was continuously rotated for 18 hours at room temperature. After hybridization the slide was washed sequentially with stirred 500 mL portions of 5X SSC, 2.5X SSC, and 1X SSC for 5 minutes each. Traces of liquid were removed by centrifugation and the slide was allowed to air-dry.

Recombinant human TNF-α (500 μg, lyophilized, from PreproTech) was taken up in 230 μL 1X PBS and dialyzed against 700 mL stirred 1X PBS at 4°C for 18 hours in a Microdialyzer unit (3,500 MWCO, Pierce). The dialyzed TNF-α was treated with EZ-Link NHS-LC-LC biotinylation reagent (20 μg, Pierce) for 2 hours at 0°C, and again dialyzed against 700 mL stirred 1X PBS at 4°C for 18 hours in a Microdialyzer unit (3,500 MWCO, Pierce). The resulting conjugate was analyzed by MALDI-TOF mass spectrometry and was found to be almost completely functionalized with a single biotin moiety.

Each of the following processes was conducted at 4°C with continuous rotation or mixing. The protein microarray surface was passivated by treatment with 1X TBS containing 0.02% Tween-20 and 0.2% BSA (200 μL) for 60 minutes. Biotinylated TNF-α (100 nM concentration made up in the passivation buffer) was contacted with the microarray for 120 minutes. The microarray was washed with 1X TBS containing 0.02% Tween-20 (3X 50 mL, 5 minutes each wash). Fluorescently labeled streptavidin (2.5 μg/mL Alexa 546-streptavidin conjugate from Molecular Probes, made up in the passivation buffer) was contacted with the microarray for 60 minutes. The microarray was washed with 1X TBS containing 0.02% Tween-20 (2X 50 mL, 5 minutes each wash) followed by a 3 minute rinse with 1X TBS. Traces of liquid were removed by centrifugation, and the slide was allowed to air-dry at room temperature.

Fluorescence laser scanning was performed with a GSI Lumonics ScanArray 5000 system using 10 μM pixel resolution and preset excitation and emission wavelengths for Alexa 546 dye. Phosphorimage analysis was performed with a Molecular Dynamics Storm system. Exposure time was 48 hours with direct contact between the microarray and the phosphor storage screen. Phosphorimage scanning was performed at the 50 μM resolution setting, and data was extracted with ImageQuant v.4.3 software.

Figures 16 and 17 are the phosphorimage and fluorescence scan, respectively, of the same array. The phosphorimage shows where the ^{10}Fn3 fusion hybridized based on the ^{35}S methionine signal. The fluorescence scan shows where the labeled TNF-α bound.

Other Embodiments

Other embodiments are within the claims.

All publications, patents, and patent applications mentioned herein are hereby incorporated by reference.

What is claimed is:

Claims

1. An array of proteins immobilized on a solid support, each of said proteins comprising a fibronectin type III domain having at least one randomized loop, at least one randomized β-sheet, or a combination thereof, and being characterized by its ability to bind to a compound that is not bound by a corresponding naturally-occurring fibronectin.

2. The array of claim 1, wherein said fibronectin type III domain is a mammalian fibronectin type III domain.

3. The array of claim 2, wherein said fibronectin type III domain is a human fibronectin type III domain.

4. The array of claim 1, wherein each of said proteins comprises the tenth module of said fibronectin type III domain (^{10}Fn3).

5. The array of claim 4, wherein each of said proteins contains one, two, or three randomized loops and wherein at least one of said loops contributes to the binding of the protein to said compound.

6. The array of claim 5, wherein at least two of said randomized loops contribute to said binding of the protein to said compound.

7. The array of claim 6, wherein at least three of said randomized loops contribute to said binding of the protein to said compound.

8. The array of claim 4, wherein said ^{10}Fn3 lacks an integrin-binding motif.

9. The array of claim 1, wherein each of said proteins lacks disulfide bonds.

10. The array of claim 1, wherein each of said proteins is a monomer or a dimer.

11. The array of claim 1, wherein each of said proteins is covalently bound to a nucleic acid.

12. The array of claim 11, wherein said nucleic acid encodes the covalently bound protein.

13. The array of claim 12, wherein said nucleic acid is RNA.

14. The array of claim 1, wherein said solid support is a chip.

15. A method for obtaining a protein which binds to a compound, said method comprising:

(a) contacting said compound with an array of candidate proteins immobilized on a solid support, each of said candidate proteins comprising a fibronectin type III domain having at least one randomized loop, one randomized β-sheet, or a combination thereof, said contacting being carried out under conditions that allow compound-protein complex formation; and

(b) obtaining, from said complex, a protein which binds to said compound.

16. A method for obtaining a compound which binds to a protein, said protein comprising a fibronectin type III domain having at least one randomized loop, at least one randomized β-sheet, or a combination thereof, said method comprising:

(a) contacting an array of proteins immobilized on a solid support with a candidate compound, each of said proteins comprising a fibronectin type III domain having at least one randomized loop, one randomized β-sheet, or a combination thereof, said contacting being carried out under conditions that allow compound-protein complex formation; and

(b) obtaining, from said complex, a compound which binds to a protein of the array.

17. The method of claim 15, said method further comprising the steps of:

(c) further randomizing a protein which binds to said compound in step (b);

(d) forming an array on a solid support with the further randomized proteins of step (c); and

(e) repeating steps (a) and (b) using, in step (a), the array of further randomized proteins as said array of candidate proteins.

18. The method of claim 16, said method further comprising the steps of:

(c) modifying the compound which binds to said protein in step (b); and

(d) repeating steps (a) and (b) using, in step (a), said further modified compound as said candidate compound.

19. The method of claim 15 or 16, wherein said solid support is a chip.

20. A method for detecting a compound in a sample, said method comprising:

(a) contacting a sample with a protein which binds to said compound and which comprises a fibronectin type III domain having at least one randomized loop, at least one randomized β-sheet, or a combination thereof, said contacting being carried out under conditions that allow compound-protein complex formation; and

(b) detecting said complex, thereby detecting said compound in said sample.

21. The method of claim 20, wherein said sample is a biological sample.

22. The method of claim 20, wherein said protein is immobilized on a solid support.

23. The method of claim 22, wherein said protein is immobilized on said solid support as part of an array.

24. The method of claim 22, wherein said solid support is a bead or chip.

25. The method of claim 15, 16 or 20, wherein said compound is a protein.

26. The method of claim 15, 16, or 20, wherein said fibronectin type III domain is a mammalian fibronectin type III domain.

27. The method of claim 26, wherein said fibronectin type III domain is a human fibronectin type III domain.

28. The method of claim 15, 16, or 20, wherein each of said proteins comprises the tenth module of said fibronectin type III domain (^{10}Fn3).

29. The method of claim 28, wherein each of said proteins contains one, two, or three, randomized loops and wherein at least one of said loops contributes to the binding of said protein to said compound.

30. The method of claim 28, wherein said ^{10}Fn3 lacks an integrin-binding motif.

31. The method of claim 15, 16, or 20, wherein each of said proteins is covalently bound to a nucleic acid.

32. The method of claim 31, wherein said nucleic acid encodes the covalently bound protein.

33. The method of claim 32, wherein said nucleic acid is RNA.

34. The method of claim 15, 16, or 20, wherein said complex or said compound is detected by radiography, fluorescence detection, mass spectroscopy, or surface plasmon resonance.

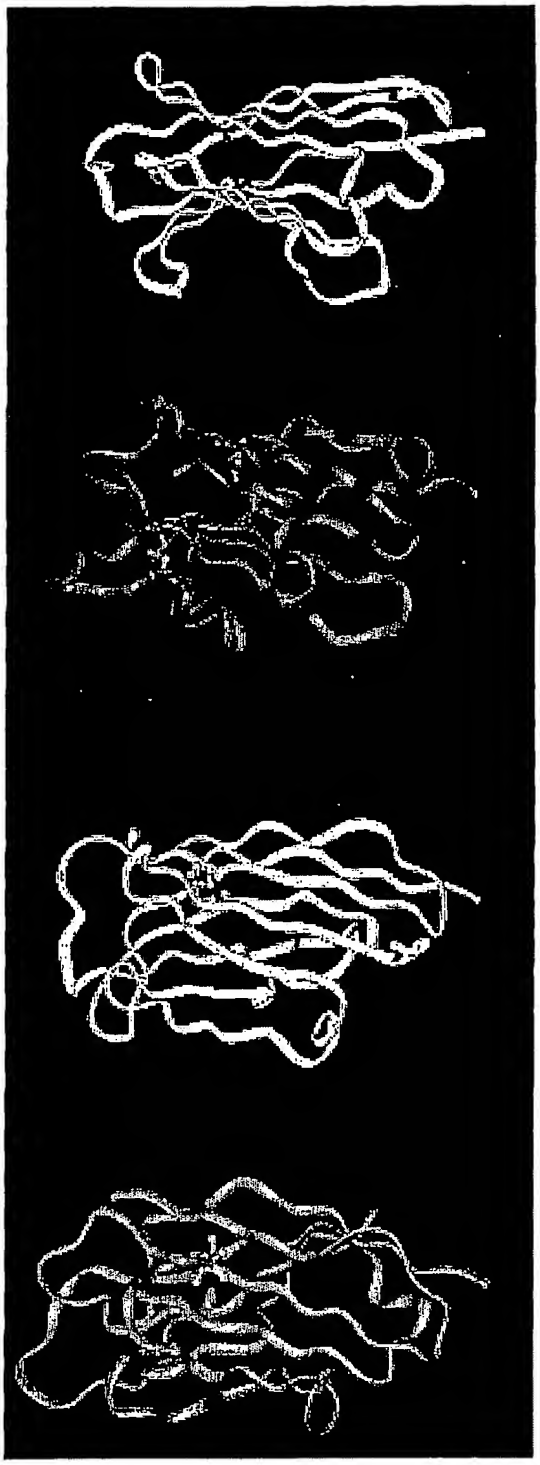


Fig. 1

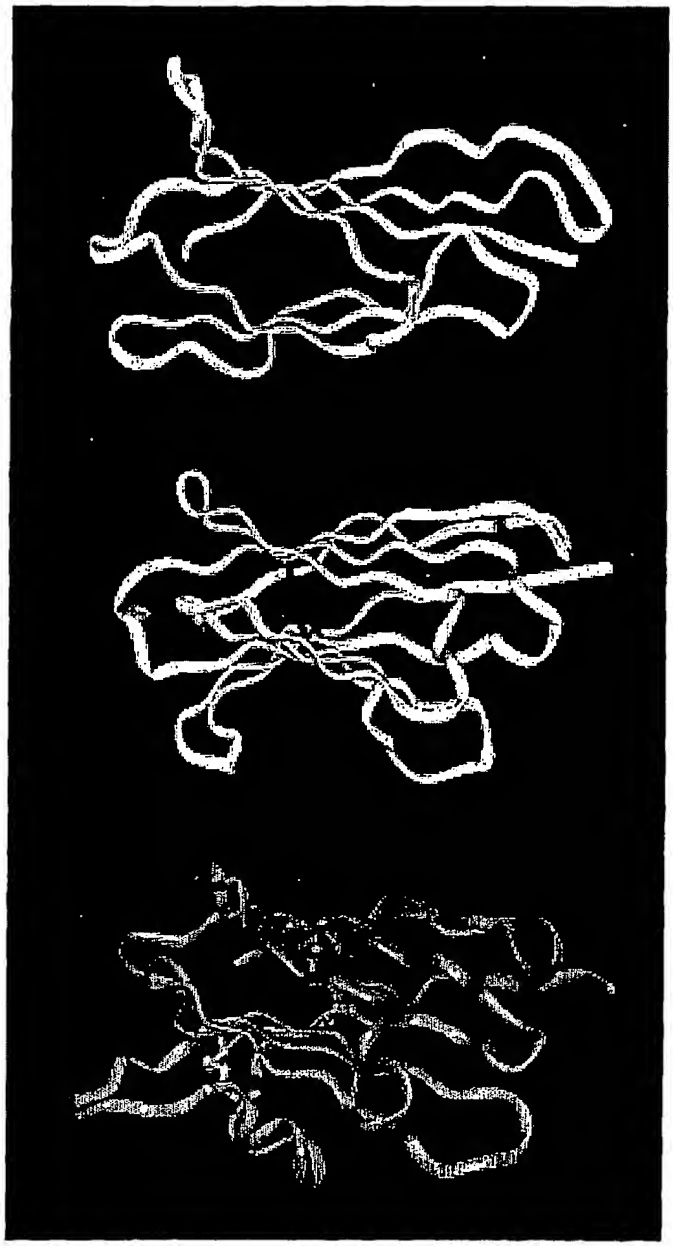


Fig. 2

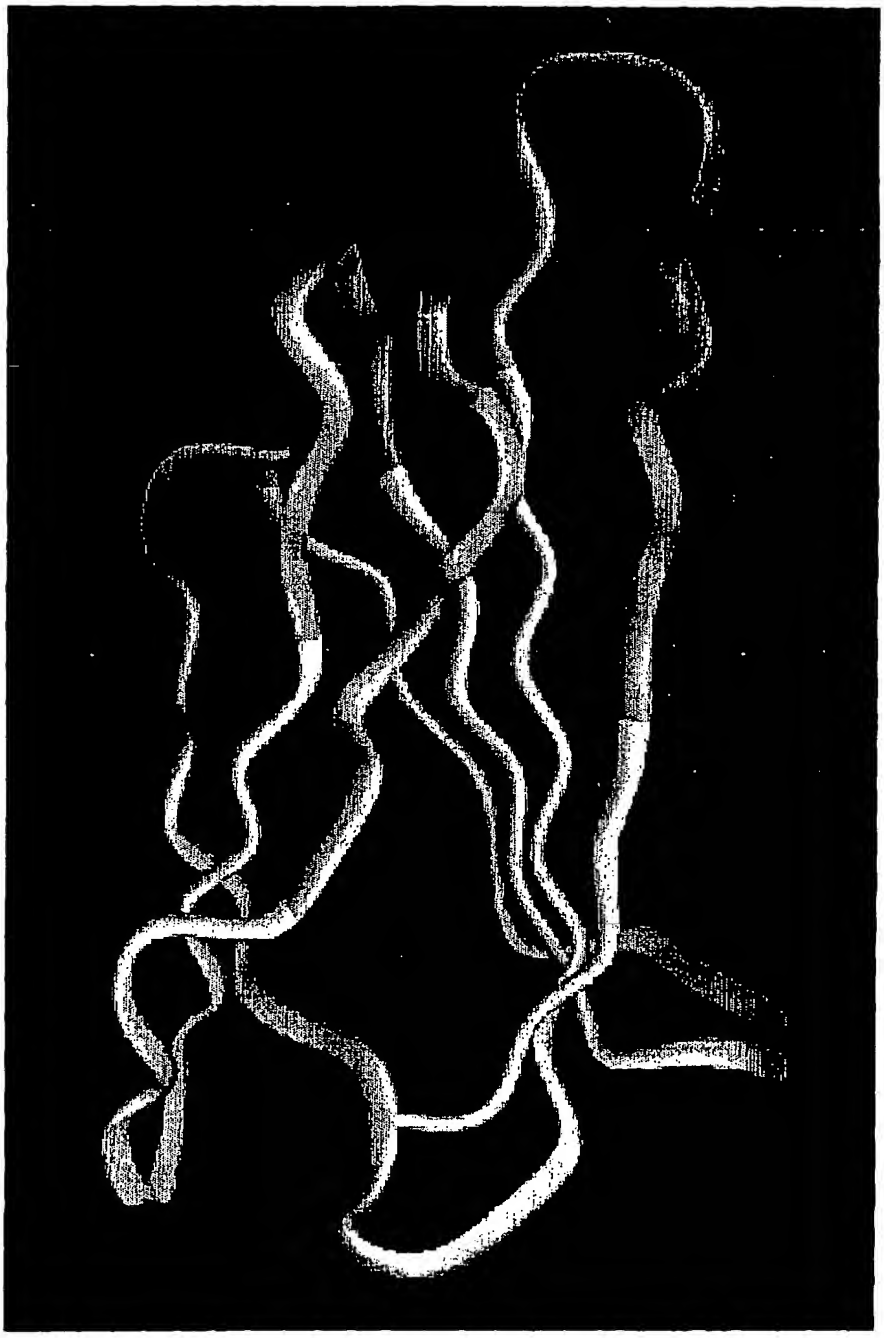


Fig. 3

```
                  1        9 10      19 20      29 30    37   38      47 48      57 58      67 68      77 78      87 88   94
Hs FND  VSDVPRD-LE VVAATPTSLL ISWDAPAVTV RYYRITYG-- ETGGNSPVQE FTVPGSKSTA TISGLKPGVD YTITVYAVTG RGDSPASSKP ISINYRT     94
Bt FN   VSDVPRD-LE VIAATPTSLL ISWDAPAVTV RYYRITYG-- ETGGSSPVQE FTVPGSKSTA TISGLKPGVD YTITVYAVTG RGDSPASSKP VSINYRT   1510
Rn FN   VSDVPRD-LE VIASTPTSLL ISWEpPAVSV RYYRITYG-- ETGGNSPVQE FTVPGSKSTA TINnIKPGAD YTITLYAVTG RGDSPASSKP VSINYQT   1611
Mm FN   VSDIPRD-LE VIASTPTSLL ISWEpPAVSV RYYRITYG-- ETGGNSPVQE FTVPGSKSTA TINnIKPGAD YTITLYAVTG RGDSPASSKP VSINYKT    712
Oc FN   VSDVPRD-LE VIASTPTSLL ISWEXPAVTV RYYRITYG-- ETppN----- ---------- ---------- ---------- ---------- -------    64
Gg FN   VSDVPRD-LE VNpTSPTSLE ISWDAPAVTV RYYRITYG-- ETGGSSPVQE FTVPGTMSrA TITGLKPGVD YTITVYAVTG RGDSPASSKP VTVTYKT    443
Xl FN   VSDVPTD-LE VTSSSPNTLT ISWEAPAVSV RYYRITYS-- QTGGhGPEKE FTVPGTSNTA TIRGLNPGVS YTITVYAVTG RGDSPASSKP LTIIHKT   1611
Cf FN   AiDAPSn-Lr FLATTPnSLL VSWQpPrArI TGYIIKYe-- kpGSpprEVV prprPGVTeA TITGLePGTE YTIQVIALKn NQKSepLIGr kKTdEL-    197
Ec FN   AiDAPSn-Lh FLATTPnSLL ISWQpPrArI TGYIIKYe-- kpGSpprEVV prphPGVTeA TITGLePGTE YTIQVIAIKn NQKSepLIGr rKTdEp-    197
Hs TC   VS-PPKD-Lv VTeVTeeTVN LAWDn-eMrV TeYLVVYTP- -THEGGlEMQ FrVPGDQTST IIQeLePGVE YFIRVFAILe NKKSipVSAr V------   686
Ss TP   VS-PPKD-Li VTeVTeeTVN LAWDn-eMrV TeYLIVYTP- -THEdGlEMQ FrVPGDQTST TIReLePGVE YFIRVFAILe NKKSipVSAr V------   686
Mm TX   MiDGPQD-Lr VVAVTPTTLD LSWlrPQAeV DrFVVSYV-- --SAGNqRVr LeVPPEADrT QLTdLMPGVE YVVTVTAERG HAVSypASIr ANTG---   889
Hs CAP  TlpVPvvSLn IYdVGPTTMH VQWQp-VGGA TGYILSYkPV kDTEpTrpKE VrLGPTVNdM QLTdLVPNTE YAVTVQAVLh dLTSepVTVr e------  1551
Oc C12  TlpVPvvSLN IYdVGPTTMH VQWQp-VGGA TGYILSYkPV kDTEpTrpKp QdVkLRdVTH ---------- ---------- ---------- -------   322
Gg C14  LalpmaSDLk LYdVShSSMR AKWnG-VAGA TGYMILYAPL TEGLAAdEKE IkIGEASTeL ELDGLLPNTE YTVTVYAMF- ---------- -------   508
Hs U1   LalpmaSDLl LYdVTenSMR VKWDA-VpGA SGYLILYAPL TEGLAGdEKE MkIGETHTdI ELSGLLPNTE YTVTVYAMFG eeASDpVTGq e------   321

var.    --- -- -  -  --       -----  -    -  ---------  - - - -    - -             - --  --  - -  - -

cons.     P  L  V     SL   W     V   Y I Y                    I  L PGVD Y ITV A   G     S      P
          M     I     TV         A   F V                    L  I  NTE   VQL     N          R
                L      M         I     L                        AS    R     E
                 F
```

CAP	Collagen alpha precursor	BOLD	identical to Hs FND
C12	Collagen type 12		
FND	Fibronectin type III domain	lower case	non-conservative substitution
FN	Fibronectin		(charge reversal, change between hydrophobic
TP	Tenascin precursor		and charged,addition or removal of P)
TC	Tenascin-C	-	position of non-conservative substitutions
U1	Undulin 1		

Bt	Bovis taurus	cow
Cf	Canis familiaris	dog
Ec	Equus caballis	horse
Ss	Sus scrofa	pig
Hs	Homo sapiens	human
Oc	Oryctolagus cuniculus	rabbit
Xl	Xenupus laevis	African clawed frog

Fig. 4

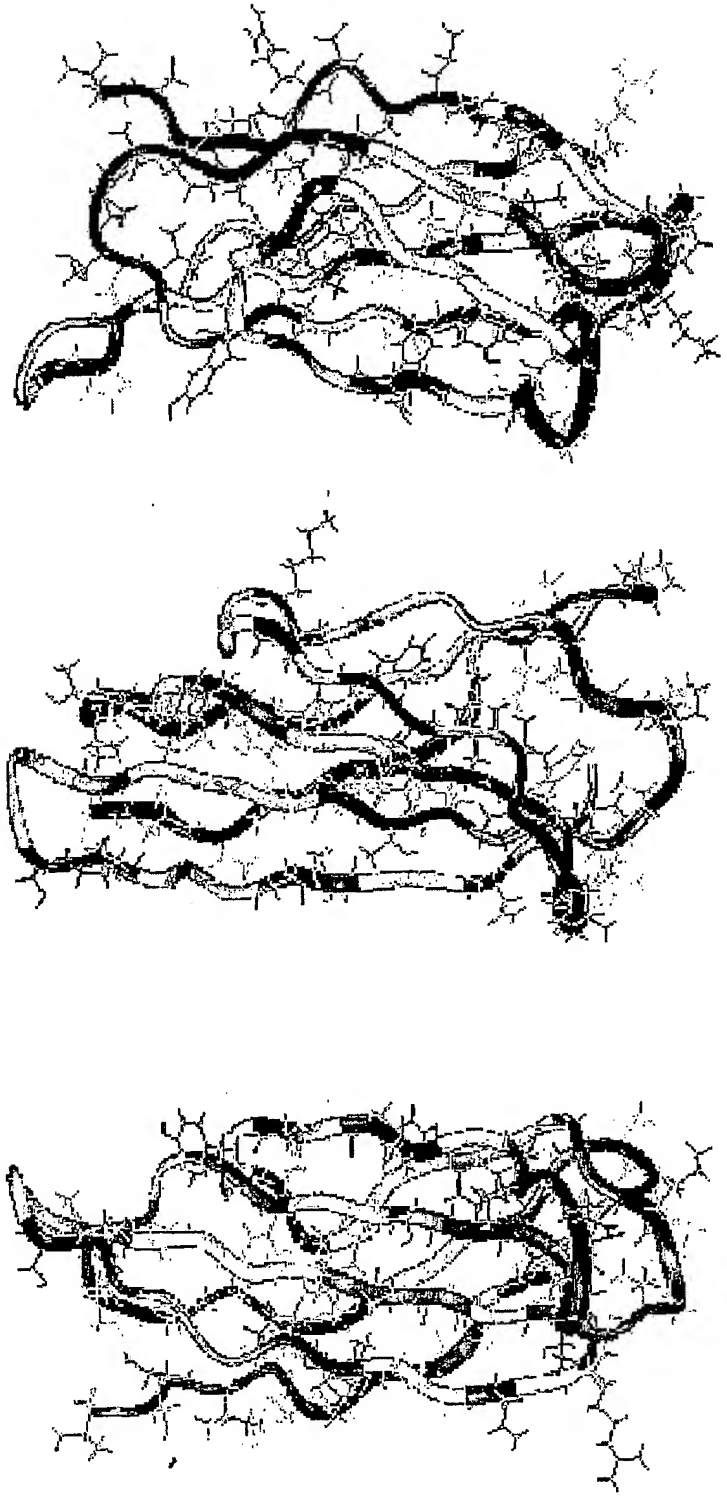


Fig. 5

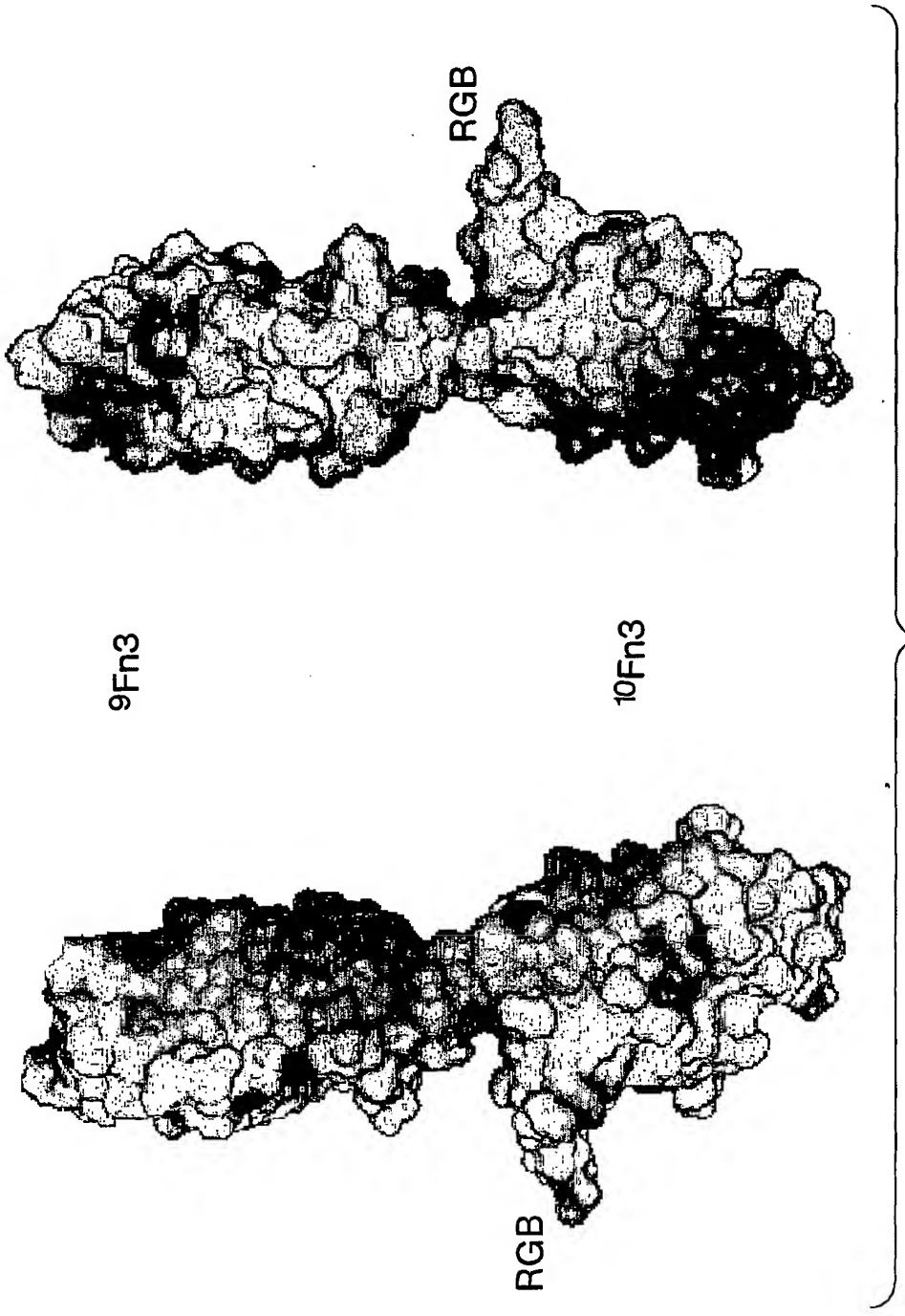


Fig. 6

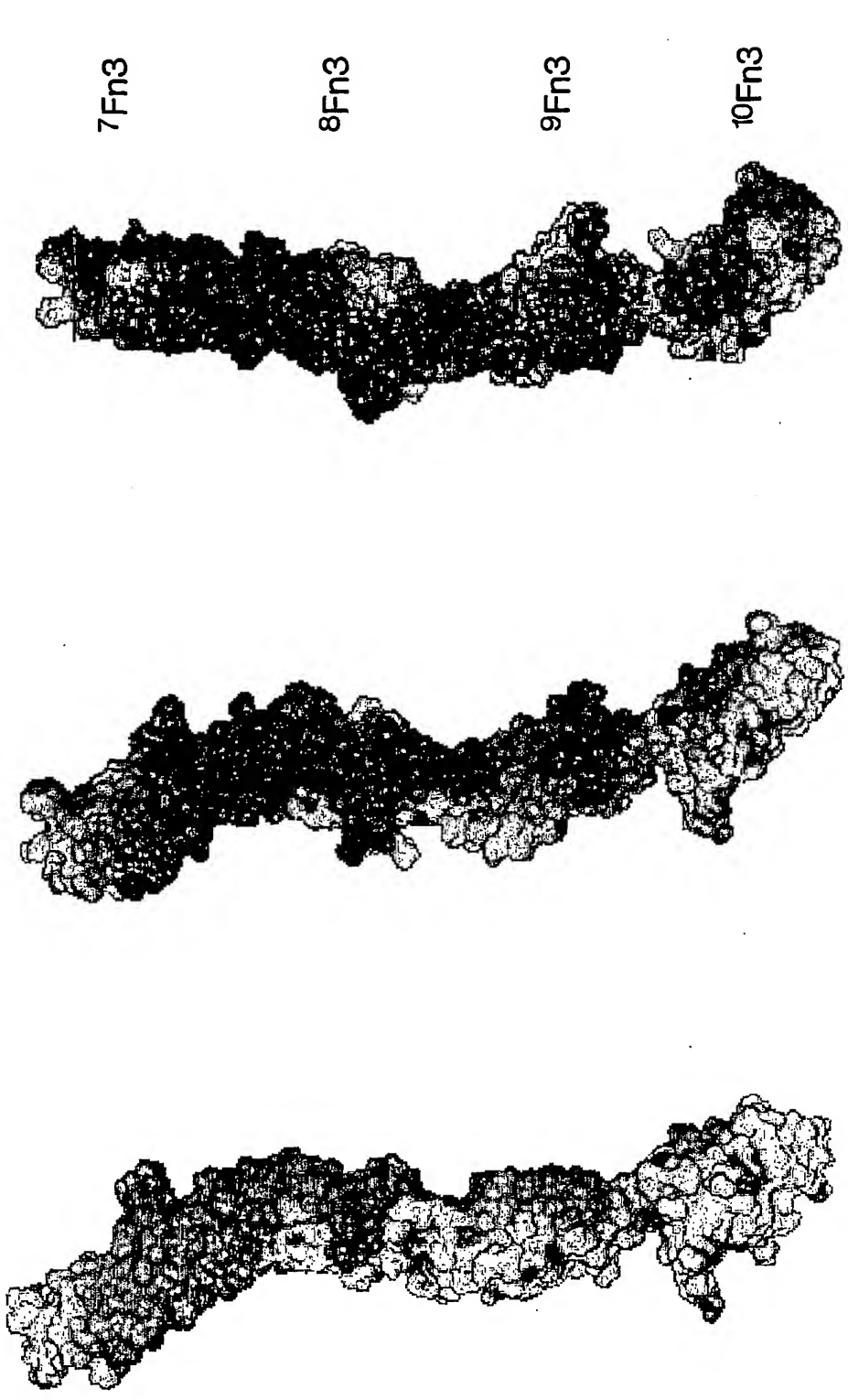


Fig. 7

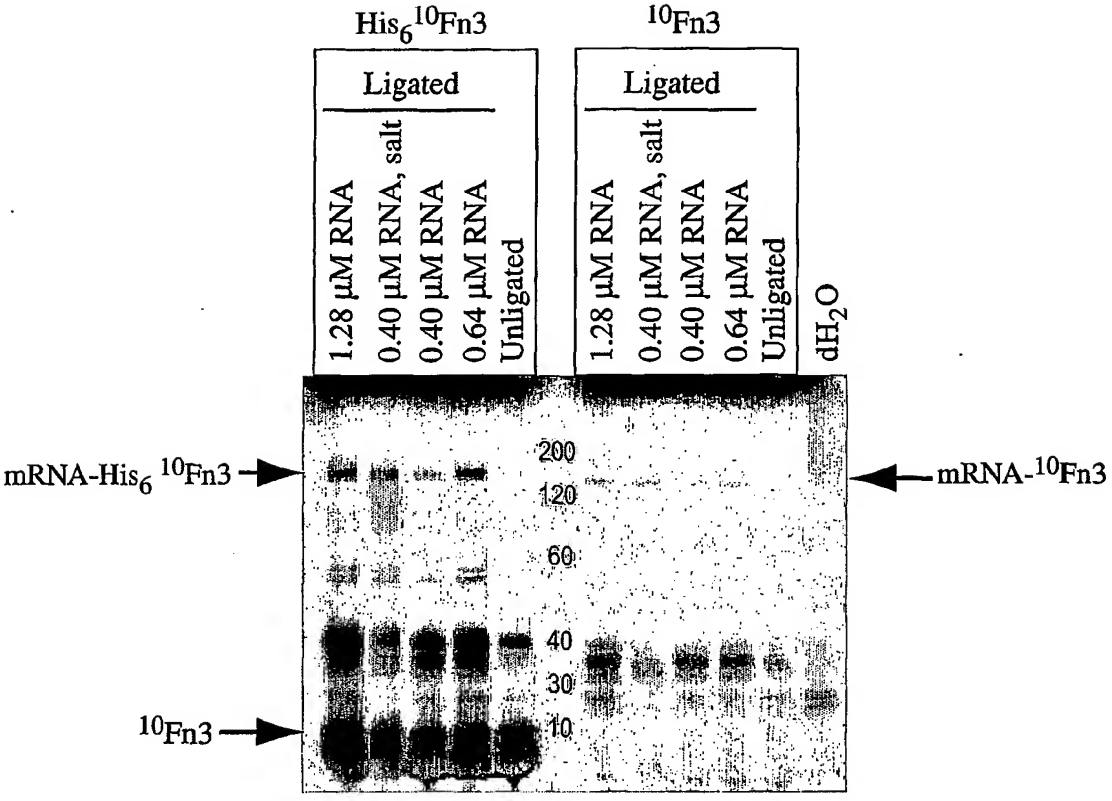

His6 10Fn3
10Fn3
Ligated
Ligated
1.28 µM RNA
0.40 µM RNA, salt
0.40 µM RNA
0.64 µM RNA
Unligated
1.28 µM RNA
0.40 µM RNA, salt
0.40 µM RNA
0.64 µM RNA
Unligated
dH2O
mRNA-His6 10Fn3
mRNA-10Fn3
10Fn3
200
120
60
40
30
10

Fig. 8

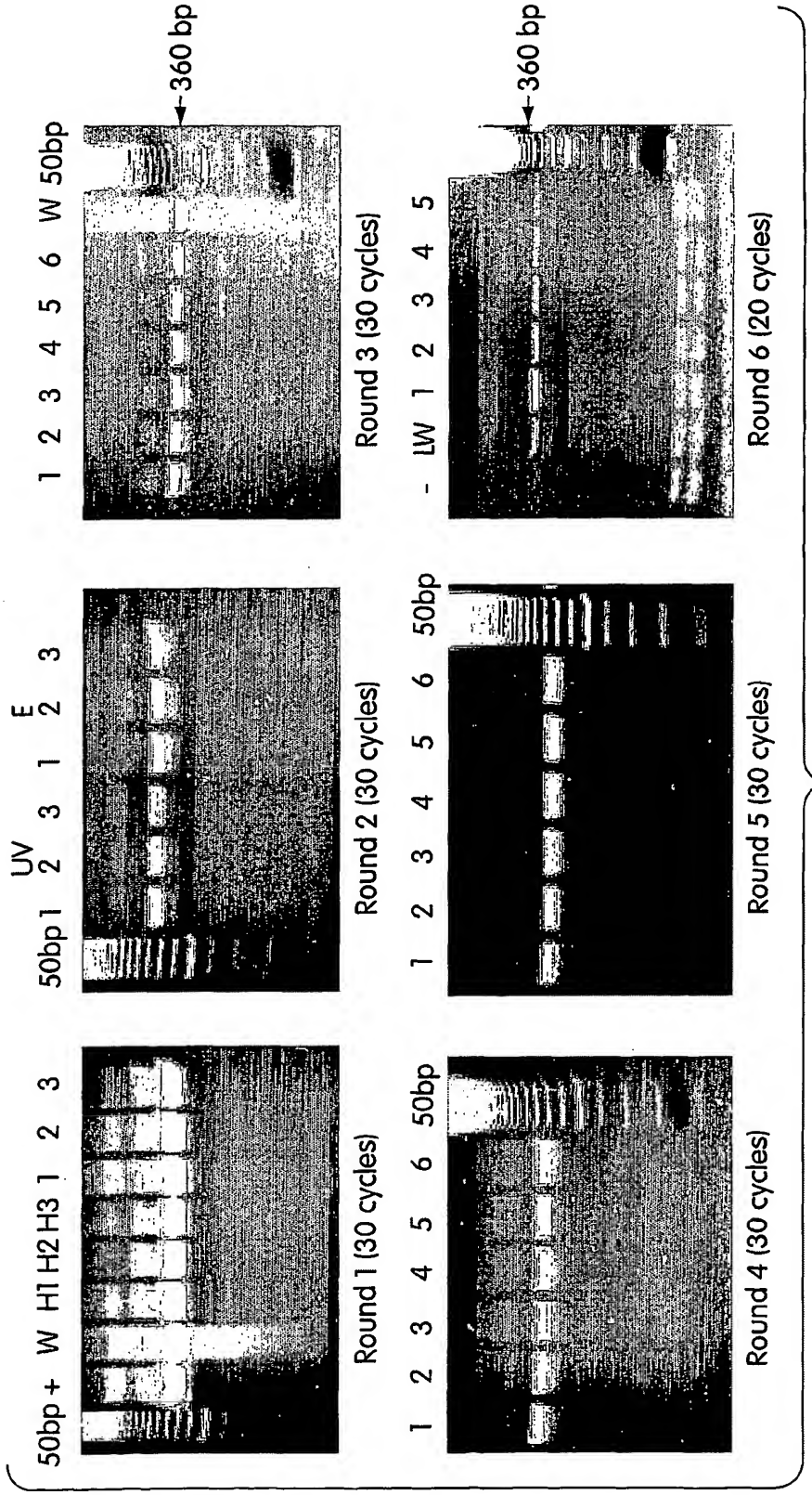

50bp + W H1 H2 H3 1 2 3
Round 1 (30 cycles)
50bp 1 2 3 UV 1 2 3 E
Round 2 (30 cycles)
1 2 3 4 5 6 W 50bp
360 bp
Round 3 (30 cycles)
1 2 3 4 5 6 50bp
Round 4 (30 cycles)
1 2 3 4 5 6 50bp
Round 5 (30 cycles)
- LW 1 2 3 4 5
360 bp
Round 6 (20 cycles)

Fig. 9

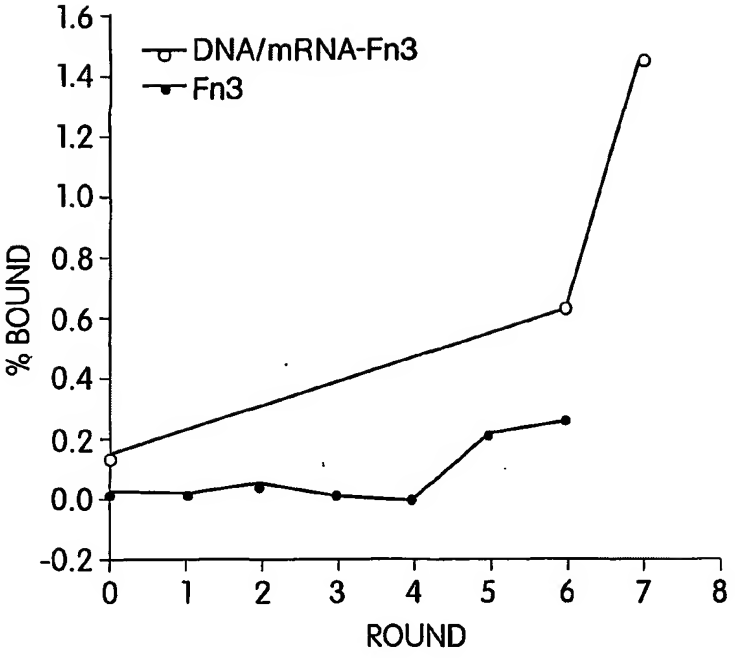

DNA/mRNA-Fn3
Fn3
% BOUND
ROUND
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
0
1
2
3
4
5
6
7
8

Fig. 10

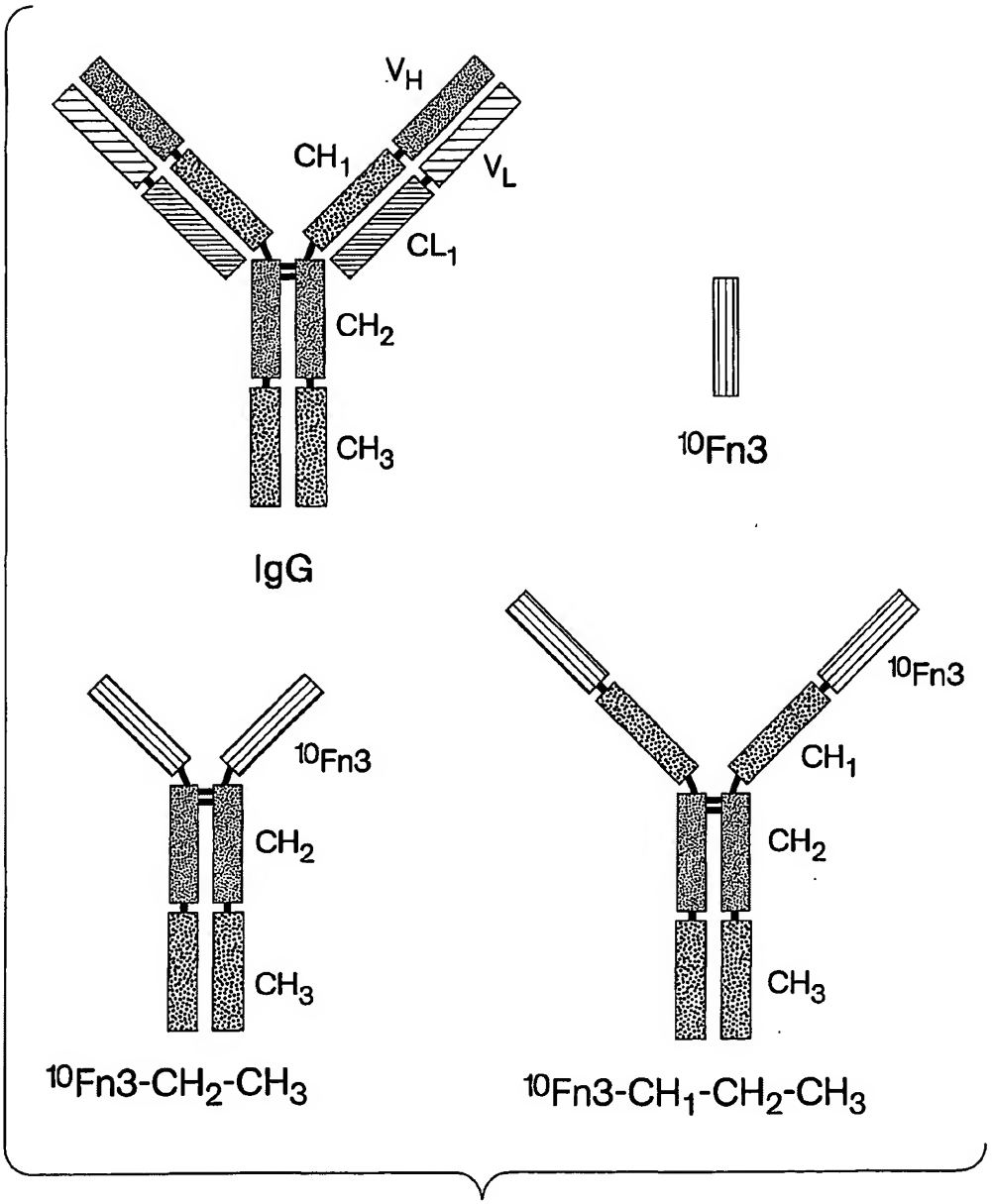

V_H
CH_1
V_L
CL_1
CH_2
CH_3
IgG
^{10}Fn3
^{10}Fn3
CH_2
CH_3
^{10}Fn3-CH_2-CH_3
^{10}Fn3
CH_1
CH_2
CH_3
^{10}Fn3-CH_1-CH_2-CH_3

Fig. 11

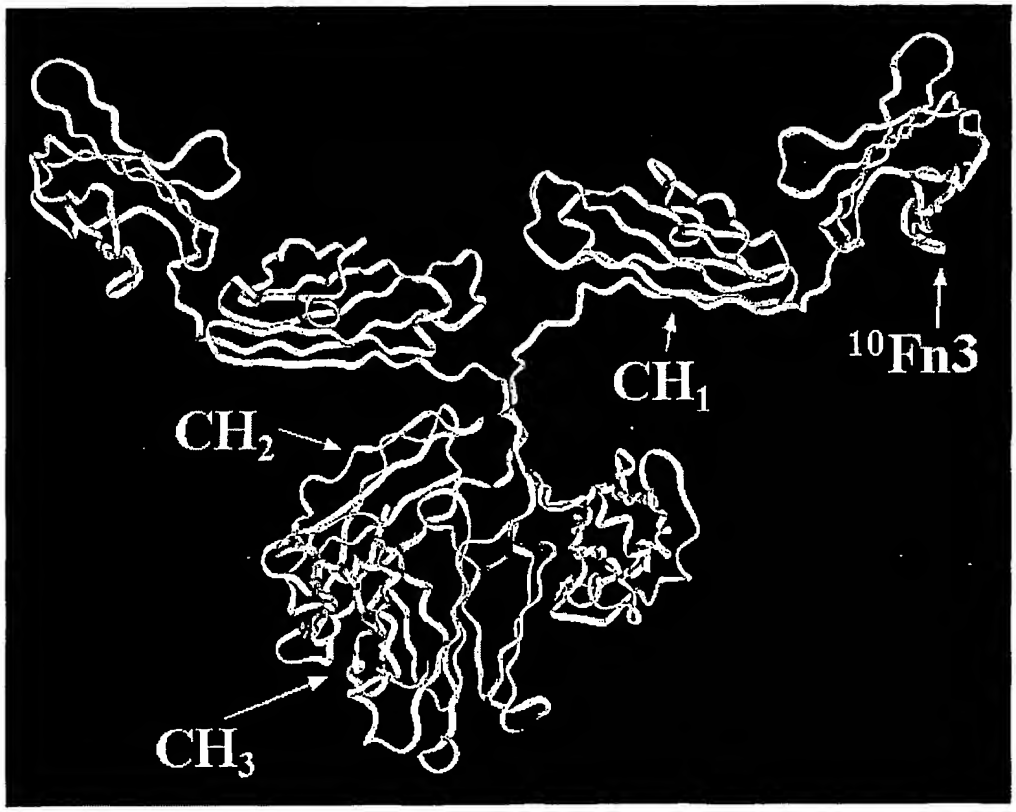


Fig. 12

FIG. 13

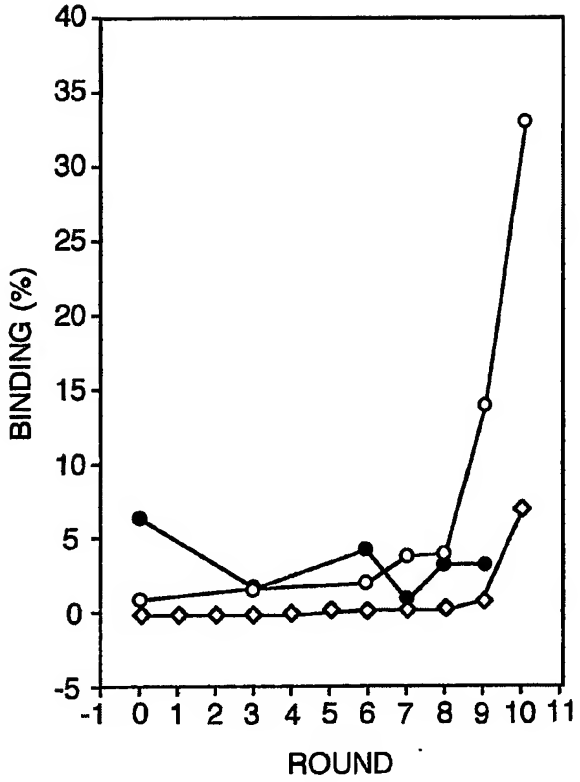


FIG. 14

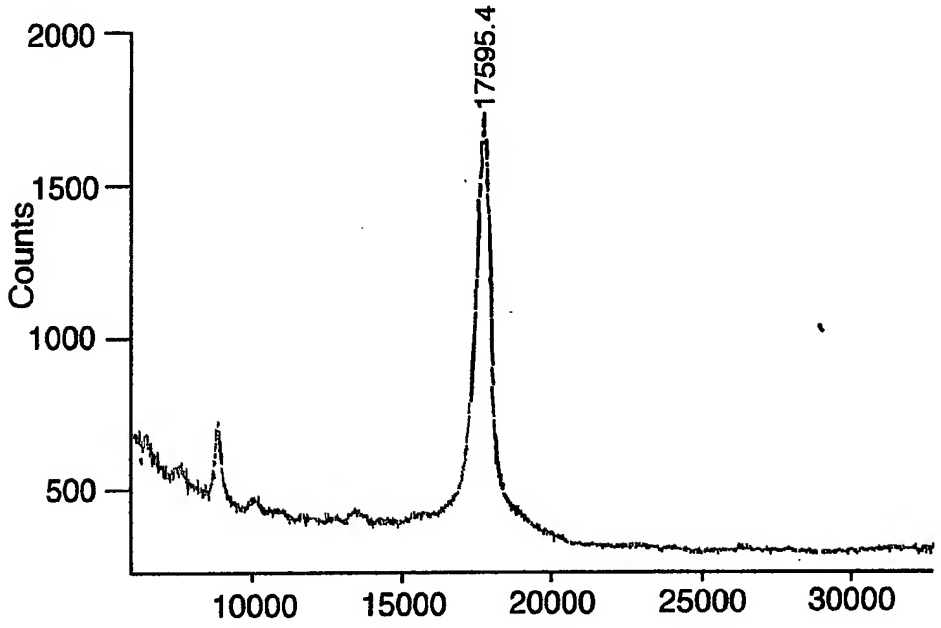

2000
1500
1000
500
Counts
17595.4
10000
15000
20000
25000
30000

FIG. 15

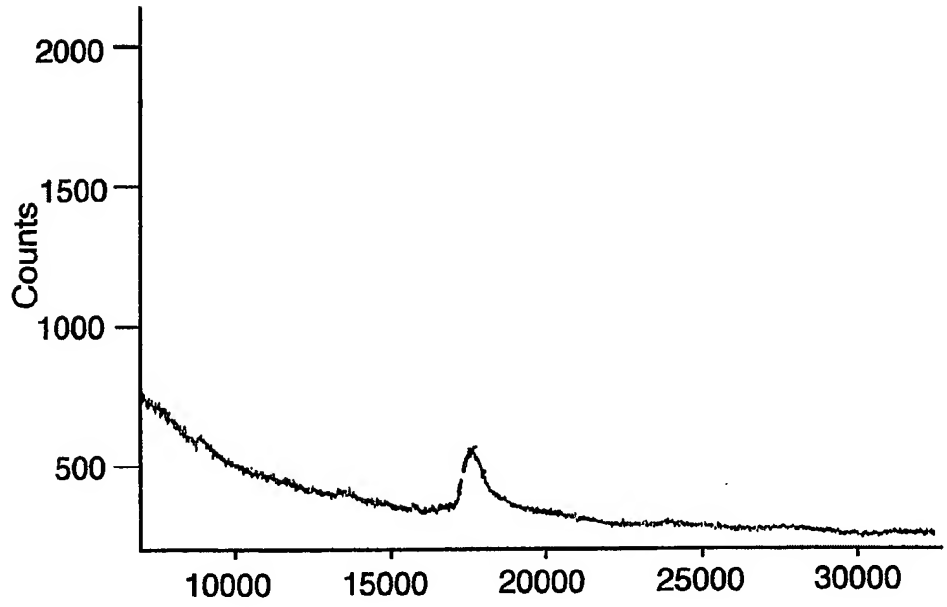

2000
1500
1000
500
Counts
10000
15000
20000
25000
30000

FIG. 16

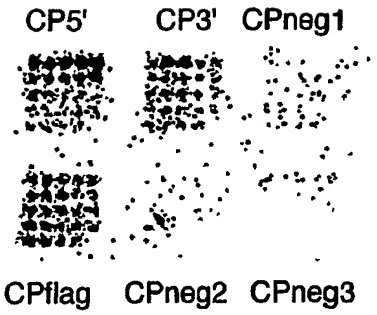
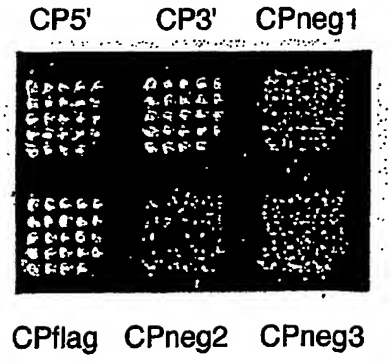

CP5'
CP3'
CPneg1
CPflag
CPneg2
CPneg3

FIG. 17


CP5'
CP3'
CPneg1
CPflag
CPneg2
CPneg3

INTERNATIONAL SEARCH REPORT

Inte onal application No.
PCT/US01/06414

A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) :C12Q 1/00, 1/70, 1/68; G01N 33/53

US CL : 435/4, 5, 6, 7.1, 69.1, 174

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/4, 5, 6, 7.1, 69.1, 174

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

Please See Extra Sheet.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X --- Y	KOIDE et al. The Fibronectin Type III Domain as a Scaffold for Novel Binding Proteins. J. Mol. Biol. 1998, Vol. 284, pp. 1141-1151, especially p. 1143, col. 2; p. 1144, Table 1; p. 1146, Col. 2.	1-6, 8-10, 15-21, 25-30 --- 11-14, 17-19, 22-24, 31-34
X ---- Y	WO 98/56915 A2 (RESEARCH CORP TECH, INC.) 17 December 1998, entire document, especially p. 20, Table; p. 26, lines 5-6; p. 27, p. 32, Ex. 1; p. 35, Ex. 5; p. 40, Ex. 10.	1-6, 8-10, 15-16, 20-23, 25-30, 34 ---- 11-14, 17-19, 24, 31-33
X, P	WO 00/34784 A1 (PHYLOS, INC.) 15 June 2000, entire document.	1-34

[X] Further documents are listed in the continuation of Box C. [ ] See patent family annex.

*	Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be of particular relevance	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E"	earlier document published on or after the international filing date	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&"	document member of the same patent family
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search	Date of mailing of the international search report
29 JUNE 2001	07 AUG 2001
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231	Authorized officer HOLLY SCHNIZER
Facsimile No. (703) 305-3230	Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet) (July 1998)*

INTERNATIONAL SEARCH REPORT

Inte ional application No.
PCT/US01/06414

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	WO 98/31700 A1 (THE GENERAL HOSPITAL CORP.) 23 July 1998, entire document.	11-14, 24, 31-34
A	NYGREN et al. Scaffolds for engineering novel binding sites in proteins. Curr. Opin. Struct. Biol. 1997, Vol. 7, pp. 463-469.	1-34

INTERNATIONAL SEARCH REPORT

Inter onal application No.
PCT/US01/06414

B. FIELDS SEARCHED
Electronic data bases consulted (Name of data base and where practicable terms used):

STN (bioscience database), EAST (all databases), search terms: fibronectin, module 10, 10fn3, scaffold, novel binding partners, combinatorial, phage display, evolv? novel binding, randomized loop.